Manuscript for **ZDM Mathematics Education**

"Profiles of in-service teachers' general pedagogical knowledge: Nature, causes and effects on teacher beliefs and instructional quality"

Authors: Caroline Nehls[1,2], Johannes König[2], Gabriele Kaiser[3, 4], Sigrid Blömeke[5]

[1] Corresponding author:
E-mail: caroline.nehls@uni-koeln.de

[2] Faculty of Human Sciences
Department of Education and Social Sciences
University of Cologne
Gronewaldstr. 2a
50931 Cologne
Germany

[3] Faculty of Education
Mathematics Education
University of Hamburg
Alsterterrasse 1
20354 Hamburg
Germany

[4] Learning Sciences Institute
Australian Catholic University
Brisbane
Australia

[5] Centre of Educational Measurement (CEMO)
University of Oslo
Gaustadalléen 30D
0373 OsloNorway

Abstract

This paper aims at identifying qualitatively different profiles of teachers' general pedagogical knowledge as a central component of their competence. We applied a mixed Rasch model (Rost 2007) to a sample of 462 in-service, mathematics and non-mathematics teachers that were tested using a short version of the TEDS-M test for general pedagogical knowledge (König et al. 2011). The analysis revealed two profiles that were characterized by (quantitative) differences in their overall GPK level as well as (qualitative) differences resulting in varying rankings of certain items' difficulties. An item-level analysis revealed that the profiles differed mainly on test items dealing with adaptivity, most notably on a set related to Bruner's modes of representation. A person-focused comparison of the profiles showed that teachers who had undergone training for teaching mathematics had a higher chance of belonging to the profile with a strength on these adaptivity items. The profiles' were validated against teacher belief and instructional quality criteria. The results showed that teachers of different GPK profiles differed significantly in their epistemological as well as teaching and learning beliefs. Moreover, teachers differed significantly in the cognitive activation level of their instruction.

1. Introduction

Over the past decades, considerable research has been conducted on (mathematics) teachers' general pedagogical knowledge (GPK) and its relationship to other facets of their professional competence such as mathematical content knowledge (MCK), mathematical pedagogical content knowledge (MPCK; König 2014; Voss et al. 2015), or self-efficacy beliefs (Dicke et al. 2015; Lauermann and König 2016). Findings have consistently emphasized the relevance of GPK for teachers' behavior in the classroom, showing that higher GPK levels are associated with higher levels of situation-specific classroom management skills (König and Kramer 2016) and students' learning gains (Brühwiler 2014; Lenske et al. 2016).

Such comparisons of averaged estimates help describe the observed relations quantitatively, but consequently tend to focus on level-differences between persons (Morin and Marsh 2014). In doing so, they treat all teachers equally and do not allow for a differentiated, qualitative analysis of their professional competences. However, teachers of the same competence level can well be assumed to differ in their profiles, i.e. the compositional patterns of their competence facets (Blömeke et al. 2008). To address this issue, a few studies have conducted person-centered

competence analyses. As one example, Kunter and Klusmann (2010) examined mathematics teachers' competence profiles by conducting a latent profile analysis of their MCK, teaching beliefs, enthusiasm, and self-regulation. The findings revealed differently composed profiles, i.e. a "problematic teacher" profile with low scores on all dimensions or a "self-regulatory teacher" profile representing a combination of slightly below-average MCK, slightly above-average constructivist beliefs and enthusiasm, and high self-regulation. The analysis further indicated an effect on teachers' instructional quality, i.e. by showing that "problematic" teachers were associated with lower cognitive activation and less constructive feedback methods than their peers from other profiles.

Patterns in teachers' GPK, however, are rarely examined under such a qualitative approach. As one example, Klemenz and König (2019) combined quantitative measures of GPK with qualitative descriptions of pre-service teachers' abilities in a proficiency level model (Hartig and Frey 2012). To do so, the authors analyzed teaching-related processes that were reflected in GPK test item attributes and related these to teachers' ability levels. Subsequently, teachers' GPK scores became interpretable in terms of how thoroughly they connected knowledge elements to solve complex problems and of how strongly they linguistically differentiated their knowledge in order to analyze and reflect their actions.

While these studies underline the value that qualitative analyses of teachers' GPK can add, their approaches have two essential limitations. On the one hand, they start from the understanding that each test item is equally difficult for all teachers. This, however, is a rather restrictive assumption, as teachers might well differ in their strengths and weaknesses even if they reach similar overall scores on a GPK test. On the other hand, many studies base the profiles on pre-defined variables such as country or item attributes. In consequence, they do not allow for an investigation of unobservable groups that might reveal profiles originating in different problem solving strategies or response tendencies (Embretson 2007).

To our knowledge, no analysis has yet taken such a qualitative look into unobservable groups of teachers' GPK patterns. Yet, such profiles could provide information on how to design learning opportunities that foster teachers' competence in accordance with their individual strengths and weaknesses (Morin and Marsh 2014). Moreover, they could provide more qualitative explanations for differences in other facets of teachers' competence than overall GPK levels do. Instructional quality, for example, has repeatedly been shown to be predicted by GPK (Brühwiler et al. 2017; König and Pflanzl 2016; Voss et al. 2014). Teachers' qualitative GPK differences (e.g. a strength/weakness in a specific dimension) could help explain why they

perform differently on certain instructional quality dimensions even if their overall GPK levels are similar.

Against this background, the present study aims at investigating the existence of qualitatively different profiles in teachers' GPK. It addresses the question whether teachers' GPK is of a heterogeneous nature and, if so, what characterizes the different profiles. For a better understanding of how to shape teachers' GPK profiles, it further investigates potential causes of such heterogeneity. Finally, it examines how the profiles relate to other competence facets such as teacher beliefs and instructional quality.

2. Theoretical framework

2.1 Defining teachers' general pedagogical knowledge

General pedagogical knowledge constitutes a central component of teachers' professional competence (Baumert and Kunter 2006; König et al. 2011; Shulman 1987). Teachers are assumed to need generic knowledge that is valid across subjects and that serves as an intellectual framework for effective classroom management and general instruction activities (Doyle 1985; Wilson et al. 1987). As such, GPK complements teachers' content knowledge and pedagogical content knowledge (Shulman 1987). While there is considerable heterogeneity in the conceptualizations of GPK in different studies (Lohse-Bossenz et al. 2018), broad consensus exists on teachers' pedagogical core demands related to classroom instruction and student learning (König et al. 2011). Thus, it is commonly regarded as knowledge of generic principles of classroom management, learners and their learning processes, teaching methods, diagnostics, educational contexts and purposes (Voss et al. 2015).

Following this understanding and Weinert's (2001) notion of competence as the successful mastering of complex professional tasks, this study agrees on a definition of GPK as a latent cognitive disposition that is required to master instruction-related tasks and includes the following: knowledge on the preparation, structuring and evaluation of lessons, on the motivation and support of students as well as their management in the classroom, on dealing with heterogeneous learning groups, and on students' assessment. Moreover, it is subdivided into different cognitive processes required to solve instructional tasks, following Anderson and Krathwohl (2001): recalling information, understanding or analyzing a concept or phenomenon, and generating ideas or options for future actions.

2.2 General pedagogical knowledge as an outcome of heterogeneous opportunities to learn

2.2.1 Pedagogical opportunities to learn in teacher education

Based on this understanding of GPK, the acquisition of GPK is regarded as a distinct objective of teacher education programs. It is often reflected in a specific part of the teacher training curriculum assigned to compulsory pedagogical classes. In many countries, such as Germany, even national standards have been established to systematically embed objectives for pedagogical competence development in teacher education programs (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland 2004).

Drawing on the concept of opportunities to learn (OTL; McDonnell 1995; Tatto et al. 2008), growing attention has recently been paid to the curricular features of teacher education programs as to analyze the conditions under which teacher competence could best be enhanced. As one example, Blömeke and Kaiser (2012) examined the profiles of OTL that pre-service mathematics teachers from the international study TEDS-M ("Teacher Education and Development Study – Learning to Teach Mathematics") had been exposed to in their teacher education. The authors identified three profiles that differed in the overall amount of pedagogical OTL as well as in the variance of pedagogical topics future teachers had been exposed to. The analysis further revealed that in most of the examined countries the majority of teachers belonged to a profile with a very broad coverage of GPK topics, indicating homogeneity within each country.

2.2.2 Germany: a case of heterogeneous pre- and in-service teacher education

Such within-country homogeneity of pedagogical OTL, however, did not apply to the German teacher education context (Blömeke and Kaiser 2012). Instead, German pre-service teachers were almost equally distributed across the OTL profiles, pointing towards heterogeneity between pedagogical teacher education programs. The standards set by the German state ministries for pre-service teachers' GPK training underlines this: the amount of pedagogical OTL during university training ranges from 30 to 90 ECTS credits depending on the school type, education level, and state (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland 2017). When asked about the focus of their pedagogical OTL, teachers additionally report very different GPK topics (Schulze-Stocker et al. 2016). This heterogeneity in OTL extends to teachers' professional development (TPD) requirements: while some states require in-service teachers to undergo 15 to 45 hours of TPD

per year, others do not require a specific amount or any evidence of training (Grothus et al. 2018).

2.2.3 Effects of opportunities to learn on teachers' GPK

This heterogeneity of pedagogical OTL becomes practically relevant, as OTL have been shown to significantly affect teachers' professional development. Blömeke and Kaiser (2012) found the different pedagogical OTL profiles to predict teachers' GPK levels. Similarly, a longitudinal study by König and Seifert (2012)showed that pre-service teachers' quantitative use of OTL correlated with their GPK levels. The findings further supported the assumption that the OTL's contents were related to the teachers' performance on the different GPK subscales.

With regard to in-service TPD in general pedagogy, empirical studies are scarce and often limited due to small samples and a lack of comparability between states and TPD features (Grothus et al. 2018). Still, meta-studies indicate that TPD measures have an effect on teachers' competence and student learning if certain conditions are met (Desimone 2009; Yoon et al. 2007). First results further support the effect of TPD on specific pedagogical facets of teachers' competence as well as their instructional practice (Alles et al. 2018; Gräsel and Trempler 2017).

2.3 Identifying profiles of general pedagogical knowledge in unobservable groups

Studies such as the one by Klemenz and König (2019) aim to identify qualitatively different profiles of teachers' GPK, but are limited to examining groups based on pre-defined item- or person-characteristics. A set of methodological approaches exists to explore latent groups and different response patterns. König et al. (2015) applied a latent class model to investigate profiles of teachers that differed in their GPK and skills to interpret specific pedagogical situations. The authors identified three classes that were best described as overall "low," "medium," and "high" level profiles, as both competence facets increased over the three classes. This illustrates a key disadvantage of the latent class model, as it assumes items not to covary (see Figure 1) and tends to create level-based profiles if this assumption is not met (Lubke and Muthén 2005; Rost 1990). For teachers' GPK, however, this assumption does not necessarily hold, as GPK test items have been found to be related by a common factor (König et al. 2011; Voss and Kunter 2013).

Figure 1. Latent class analysis model
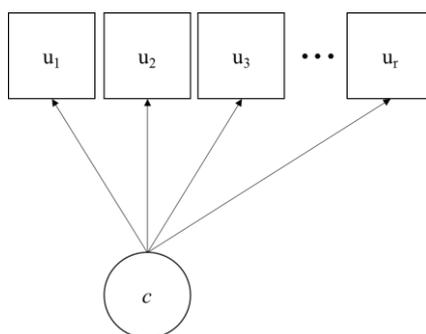
In contrast, differential item functioning (DIF) models take into account the common factor underlying the items while (as indicated by "*f*" in Figure 2) additionally allowing for qualitatively different response patterns between manifest groups (as indicated by "*c*"). Lohse-Bossenz et al. (2018) applied DIF analysis to GPK profiles to compare pre-service teachers of different teaching subjects. The findings indicated that teachers of mathematics and natural sciences had a significantly higher probability of solving "diagnostics" and "assessment" items than teachers of social sciences or languages. While the findings provide valuable information on the role of subject-specific teacher education in developing GPK, they also underline how DIF analysis limits the detection of profiles to pre-defined groups such as gender, country or teaching subject.

Figure 2. DIF analysis model
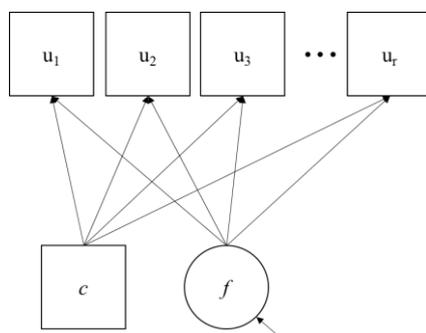
The limits of latent class and DIF analysis are addressed by factor mixture models (Lubke and Muthén 2005; Lubke and Neale 2008). As in DIF analysis, these account for the latent factor structure by holding it equal across groups (indicated by "*f*" in Figure 3), but allow for response patterns and item difficulties to differ. As in latent class analysis, they take an exploratory look in order to identify previously unknown groups (indicated by the encircled "*c*"). Factor mixture models have been used to detect profiles in students' performance on reading comprehension and natural science tests (Baghaei and Carstensen 2013; Kleickmann et al. 2011), but are rarely applied to teachers. As one exception, Morin and Marsh (2014) used the model to identify

profiles in university teachers' instructional quality as indicated by nine sub-dimensions. The analysis revealed five profiles, of which two could be described as generally poor or good. The remaining profiles differed in their levels and strengths, representing an overall average but goal-oriented group, an affective-relationally focused group, and a mastery-oriented group.

Figure 3. Factor mixture analysis model
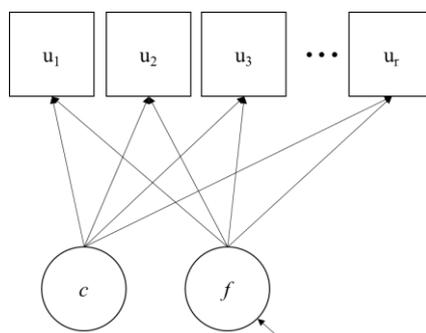
2.4 Relevance of general pedagogical knowledge for teacher beliefs and instructional quality

Several studies have underlined the relationship between teachers' professional knowledge and other competence facets, such as their beliefs, as well as the relationship between knowledge and instructional practice. Teachers' MCK and MPCK have turned out to be negatively correlated with static beliefs on the nature of knowledge and knowing as a set of strict rules and procedures to be learned and positively correlated with dynamic views of knowledge as an active process of individual enquiries and diverse problem solving approaches (Blömeke 2012; Tatto et al. 2008). Similarly, MPCK has been found to affect the development of beliefs about teaching and learning by reducing transmission-oriented beliefs and increasing constructivist views of learning as an active self-directed construction process (Blömeke et al. 2014). Even though such relations have not yet been established for GPK, they can well be assumed for subject-independent aspects of these beliefs, as student-orientation and individualized, differentiated learning are core aspects of GPK on adaptivity.

Another facet that GPK has been shown to affect, is GPK-related instructional quality including classroom management, student support and cognitive activation (Praetorius et al. 2018). Classroom management refers to a quality-oriented use of learning time and effectively preventing or dealing with disorders. Cognitive activation encompasses challenging tasks and instructional strategies that activate students' high level thinking, as part of problem induction processes (Minnameier et al. 2015). Student support relates to socio-emotional aspects such as a positive learning climate, but also includes cognitive aspects of differentiation in problem *solving* processes, such as scaffolding (Minnameier et al. 2015).

According to recent findings, higher GPK levels are associated with teaching methods/teacher clarity and teacher-student-relationships (König and Pflanzl 2016), showing medium-term effects on classroom management and constructive learner support (Lohse-Bossenz et al. 2015; Voss et al. 2014). Findings on cognitive activation are not fully consistent. While some point towards practically relevant associations in some studies (König, Kaiser et al. under review; Voss et al. 2011), others do not (Depaepe and König 2018; Voss et al. 2014) – which might indicate links to specific GPK aspects and a strong link to subject-specific knowledge facets at the same time.

3. Research questions

Our study addresses four research questions. The first relates to the existence of qualitatively different knowledge profiles.

> (RQ 1)   Do teachers differ in their overall GPK level as well as in the shape of their GPK profiles?

Taking into account the quantitative and qualitative variation of OTL within the German pre- and in-service teacher education system as well as their effects on GPK, it seems unlikely for teachers' GPK to be of uniform shape and differ by level only. Instead, we expect to find GPK profiles with level and shape differences. Given our explorative approach of looking into unobservable groups, we refrain from hypothesizing a specific number of profiles.
The second question addresses the nature of such profiles.

> (RQ 2)   If teachers' GPK is indeed of heterogeneous nature, which characteristics do the profiles have in common and what distinguishes them from one another?

Considering the diversity of German regulations for teacher education as well as the differences in teachers' use of pedagogical OTL, we expect profiles with different strengths and weaknesses on items of certain GPK sub-dimensions. Findings on teachers' cognitive processes in the area of GPK further suggest that teachers might perform differently, depending on the complexity level of the cognitive processes they require. Thus, in terms of quantitative differences, profiles of higher overall GPK levels could be expected to more likely solve highly complex items. In terms of qualitative differences, however, we expect the ranking of the item groups to remain equal between the profiles, so that highly complex items prove more difficult than low complexity items within each profile.
The third research question regards the source of profile heterogeneity.

(RQ 3)   What individual background characteristics can explain the teachers' likelihood of belonging to a specific GPK profile?

To this date, research on teachers' profiles is mostly focused on observable groups. Yet, considering results from DIF and OTL analyses, we expect structural characteristics of the teachers' individual education backgrounds to predict their assignment to the different profiles. Finally, our fourth question focuses on the profiles' association with other measures of teacher competence, as it would provide first evidence for the profiles' validity.

(RQ 4)   Are the GPK profiles related to differences in other facets of teachers' competence, such as beliefs or instructional quality?

As several studies have underlined the influence of teachers' knowledge on their beliefs, we expect teachers' GPK profiles to be associated with different epistemological beliefs and beliefs on teaching and learning. In light of recent findings on the effect of GPK on instructional quality, we further expect profile differences to show on the GPK-related dimensions of instructional quality. Which of the dimensions are most likely affected and how, will depend on their specific levels and shapes.

4.   Method

4.1   Sample and data collection

To address our research questions we used a sample of in-service teachers from Germany. It consisted of 462 mathematics and non-mathematics teachers taken from three studies: the TEDS-Follow Up, CME- and TEDS-Validate study (König et al. 2015; König 2015). Teachers had completed teacher training for primary (8%) and secondary school level (92%), and had an average teaching experience of about 11 years ($SD = 11.0$; min $= 0.5$; max $= 41.0$). Information on the specific subjects teachers had studied was missing for a substantial part of the sample, as it had not been collected in the CME study. However, the information on the school type teachers had been trained for could serve as an alternative indicator for teachers' training in mathematics teaching, as the Ministry of Education in North Rhine-Westphalia (where the CME study took place) requires all primary school teachers to either make mathematics their core subject or take basic courses in mathematics teaching (Ministerium für Schule, Jugend und Kinder des Landes Nordrhein-Westfalen 2004). As a result, our sample can be described as follows: 8% ($n = 35$) had undergone training for teaching mathematics at primary school level and 74% ($n = 343$) for mathematics at secondary school level, while 18% ($n = 81$) had been trained for teaching other subjects.

4.2 Test instruments

4.2.1 General pedagogical knowledge

Teachers' GPK was assessed using the standardized paper-and-pencil test from the TEDS-M study (König et al. 2011). Following our understanding of GPK, the test was conceptualized in a task-based way. It focused on core instructional tasks, covering the four dimensions "structure," "motivation/classroom management," "adaptivity," and "assessment." These were combined with different cognitive demands and formed a test design matrix as shown in Table 1.

Table 1. Design matrix of the TEDS-M GPK test (König et al. 2011)

Teachers' professional tasks in the instructional process		Cognitive demands		
Sub-dimensions	Operationalized topics	recall	understand/ analyze	generate
Structure	- structuring of learning objectives - lesson planning and structuring the lesson process - lesson evaluation			
Motivation/ classroom management	- strategies to prevent and counteract interferences - effective use of allocated time/ routines - achievement motivation - strategies to motivate single students/ the whole group			
Adaptivity	- strategies of differentiation - use of a wide range of teaching methods			
Assessment	- assessment types and functions - evaluation criteria - teacher expectation effects			

Due to data collection constraints an abbreviated version of the test was implemented, reducing test length to 20 minutes. Both the extended and the short version of the instrument have proven effective in measuring teachers' GPK in several studies, providing evidence for its construct, curricular, and predictive validity (for an overview, see König 2014; Voss et al. 2015). The short version comprised a total of 15 tasks including 47 items, in multiple-choice as well as open-response formats. These covered the same cognitive demands and content dimensions as the extended test version (see Table 2). Figure 4 and Figure 5 exemplify a multiple-choice and an open-response task that require teachers to analyze examples of student motivation and generate questions for lesson analysis.

Figure 4. Exemplary task of the TEDS-M GPK test (PK01C), combining the sub-dimension "motivation" and the cognitive process of "analyzing" (König et al. 2011)

Which of the following cases represents an example of intrinsic motivation, and which represents an example of extrinsic motivation?

Check <u>one</u> box in each <u>row</u>.

A student learns before a test in mathematics, because he/she...	intrinsic motivation	extrinsic motivation
A. expects a reward for a good grade.	$\square_1$	$\square_2$
B. wants to avoid the consequences of a bad grade.	$\square_1$	$\square_2$
C. is interested in problems of mathematics.	$\square_1$	$\square_2$
D. does not want to disappoint his/her parents.	$\square_1$	$\square_2$
E. wants to maintain his/her relative rank in the class.	$\square_1$	$\square_2$

Figure 5. Exemplary task of the TEDS-M GPK test (PK39), combining the sub-dimension "structuring" and the cognitive process of "generating" (König et al. 2011)

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

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?*

Figure 6. Exemplary task of the TEDS-M GPK test (PK93), combining the sub-dimension "classroom management" and the cognitive process of "understanding/analyzing" (König et al. 2011)

When the teacher poses a question in class, one student repeatedly raises his hand.
Is this an example of "operant conditioning"?

Check <u>one</u> box only.

A. Yes, this student exhibits a particular behavior and receives a positive consequence.	$\square$
B. Yes, his behavior is caused by external intervention.	$\square$
C. No, his behavior is caused by classical conditioning.	$\square$
D. No, this behavior occurs spontaneously and is not learned.	$\square$

Figure 7. Exemplary task of the TEDS-M GPK test (PK38), combining the sub-dimension "assessment" and the cognitive process of "generating" (König et al. 2011)

> 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.

Figure 8. Exemplary task of the TEDS-M GPK test (PK24B), combining the sub-dimension "adaptivity" and the cognitive process of "understanding/analyzing" (König et al. 2011)

> Which advantages do "Learning stations" offer in class?
>
> Name three different advantages of this teaching and learning method.

For each sub-sample, open-response items were coded independently by two raters who were specifically trained in using the coding scheme by König and Blömeke (2010). As a measure of consensus Cohen's Kappa was estimated, showing good results ($\kappa \geq .79$).

The items were further recoded according to the approach outlined by Klemenz and König (2019), which has proven useful for qualitatively describing GPK-related processes of pre-service as well as in-service teachers (Nehls et al. under review). Through this recoding approach items that represented partial steps of an overarching, complex process were merged into one, allowing for a qualitative description of the items as representing cognitively high or low complexity (Eye 1999). This process led to 32 final test items, of which 21 could be classified as requiring low and 11 as requiring high complexity (see Table 2).

Table 2. Tasks per sub-dimension and cognitive complexity level in the short version of the TEDS-M GPK test

Sub-dimension	Low complexity	High complexity
Structure	PK82	PK33, PK32, PK39
Motivation/ classroom management	PK01, PK09	PK40A, PK67
Adaptivity	PK20B, PK20C, PK26	PK22A
Assessment	PK12	PK07, PK15

4.2.2   Teacher beliefs

Beliefs were captured using an instrument from the TEDS-M study that required teachers to indicate their agreement with 24 items on 6-point Likert scales (from 1 "do not agree at all" to 6 "fully agree"). The items covered student- and teacher-oriented teaching and learning views, as well as static and dynamic views on the nature of knowledge in 6 statements each (see Blömeke et al. under review in this issue). Reliability was acceptable, with Cronbach's α ranging between .64 and .73.

4.2.3 Instructional quality

Instructional quality was evaluated using an observational protocol specifically developed for live ratings by external observers (Jentsch et al. under review). The protocol covered three GPK-related dimensions of instructional quality: "classroom management" (3 items), "student support" (4 items), and "cognitive activation" (7 items), all rated on 4-point Likert scales ranging from 1 ('does not apply at all') to 4 ('does fully apply'). Reliability was good for all scales ($.73 \leq \alpha \leq .87$).

4.3 Data analysis

To investigate RQ 1, we modelled teachers' GPK using a one-dimensional Rasch model and several mixed Rasch models (MRM; Rost 1990) with increasing numbers of classes. As a specific type of factor mixture model, the MRM enables the extraction of qualitatively different classes while, within each group, holding on to the one-parameter logistic model (Ayala and Santiago 2017). The analyses were applied using the software package M*plus* 7.4 (Muthén and Muthén 1998-2012). The models were compared using the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the sample-size adjusted BIC (aBIC). A lower value on these criteria served as an indicator of which model best represented the data.

To identify the number of substantially different profiles, a number of statistical tests and indices are available. Likelihood ratio tests provide guidance by comparing a *k*-class solution to a *(k - 1)*-class solution and indicating that the *(k - 1)*-class model should be rejected when their *p*-value is significant. Among these, the adjusted Likelihood Ratio Test (aLRT; Lo et al. 2001) and the bootstrap Likelihood Ratio Test (bLRT) are most commonly recommended (Nylund et al. 2007). Further, the entropy provides information on the accuracy by which a model classifies persons into the various classes, varying from 0 to 1 and indicating fewer classification errors with higher values (Ramaswamy et al. 1993).

We addressed RQ 2 by studying the GPK profiles on an item-level. The teachers' item response probabilities were compared within and between profiles, in order to detect what items most clearly distinguished them. These were then analyzed with regard to content and format. In order to assess whether the profiles' differences in item difficulties depended on these item features, Mann-Whitney-U and Kruskal-Wallis tests were performed.

Based on the commonalities of these features, we identified an OTL variable that could help explain why some teachers' had more difficulties on certain items than others (RQ 3). Whether this variable was indeed related to teachers' class membership, was examined in a Pearson Chi-

square and a Cramer's *V* test. To assess whether the profiles were related to teachers' beliefs and instructional quality (RQ 4), *t*-tests for independent samples were conducted.

4.4 Identifying knowledge profiles

To test whether teachers showed GPK profiles that differed in level and shape (RQ 1), we fitted a single-class Rasch model as well as 2- to 3-class MRMs to the data. Table 3 displays the fit indices that resulted from the estimated models. The three information criteria were lower (better) for both MRMs than for the single-class Rasch model, underlining the existence of at least two qualitatively different GPK profiles. Among the different MRMs the fit indices pointed towards the 2-class model as the best solution. While entropy was highest for the 3-class solution, the aLRT showed no significant result, indicating that the 3-class model did not fit the data significantly better than the 2-class model. This was further supported by the BIC, which is the most recommended criterion for selecting factor mixture models (Nylund et al. 2007) and was lowest for the 2-class model.

Table 3. Fit indices for one- to four-class models

Model	#par	AIC	BIC	aBIC	aLRT	bLRT	entropy	class sizes
Rasch Model								
1 factor, 1 class	33	12,025.83	12,162.30	12,057.57	n/a	n/a	n/a	462
MRM								
1 factor, 2 classes	67	11,566.68	11,843.76	11,631.12	< .001	< .001	0.88	271; 191
1 factor, 3 classes	101	11,446.95	11,864.64	11,544.09	*n.s.*	< .001	0.92	21; 270; 171

Note: #par = number of parameters; logL = Loglikelihood; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; aBIC = adjusted BIC; aLRT = adjusted likelihood ratio test statistic; bLRT = bootstrap LRT statistic.

4.5 Analyzing profile differences

In a second step, we analyzed the two class-solution with regard to the quantitative and qualitative differences (RQ 2). The MRM results showed that teachers assigned to class 1 clustered around a higher GPK factor score ($M_{Class1} = 0.293$, $S.E._{Class1} = 0.108$) than those assigned to class 2[1]. The difference represented a small but significant effect, $z = 2.70$, $p = .007$, Cohen's $d = .23$. Despite the overall GPK level difference, the classes could not be merely characterized as "high" or "low" level groups, since class 2 not only comprised teachers on the lower end of the GPK scale ($min_{Class1} = -1.52$; $min_{Class2} = -2.56$), but also teachers with a GPK level above that of class 1 teachers ($max_{Class1} = 1.15$; $max_{Class2} = 1.40$).

---

[1] The mean of class 2 is set to zero by default in M*plus*, so that the mean of class 1 represents the difference of latent factor means between the two classes.

On an item-level, the profiles further showed a mix of quantitative differences (between correct response probabilities) and qualitative differences (in the items' difficulty rankings), as can be seen in Explaining profile membership

. To illustrate the possible combinations, a few items shall be highlighted, without going into the details of their meanings at this point. For example, class 2 had a 17 percentage points higher probability of solving PK09SB, but an almost equal relative difficulty doing so, as the item was among the three most difficult items in both classes. In contrast, the classes had an almost equal probability of solving PK20C01 (C1: 10%, C2: 8%), but large differences in their relative difficulties, since the item ranked first in class 2 and 11th for class 1 (1 being the easiest, 32 the hardest). Few items, such as PK26S01B, PK26S02B, and PK26S03B, showed differences in probability *and* rank at the same time: the probability differences amounted to over 80 percentage points, making the items some of the easiest for class 1 and some of the hardest for class 2.

Figure 9. Profiles of conditional item response probabilities in the two-class MRM solution
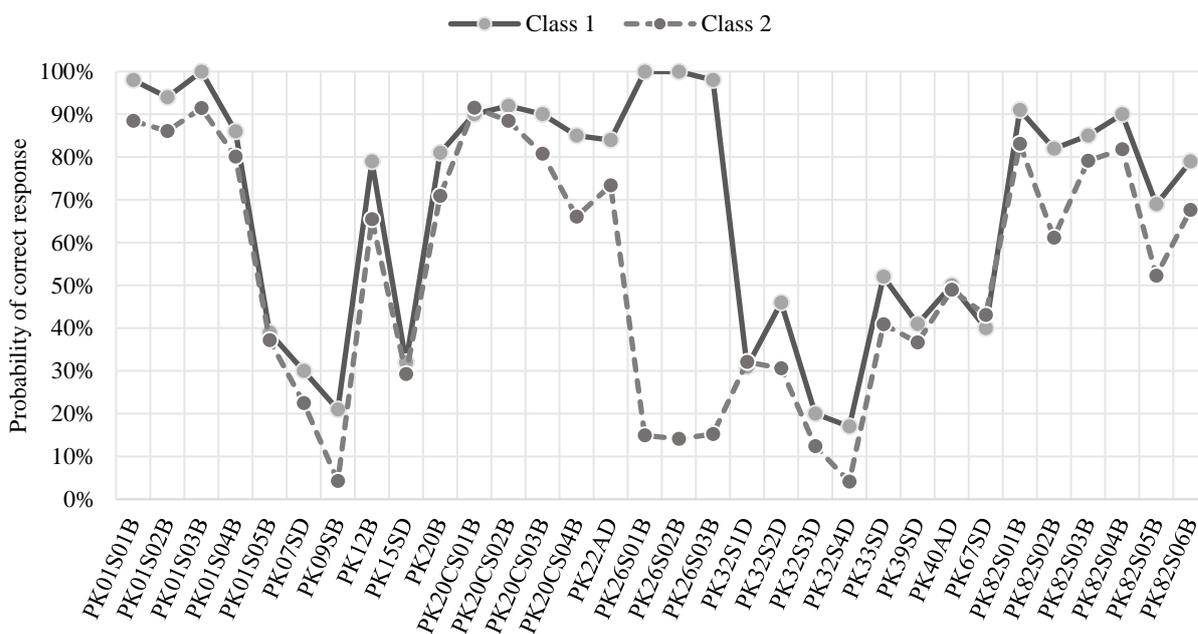
To test whether the profile (dis-)similarities could be explained by item features, the items were grouped according to the cognitive complexity level and GPK sub-dimension they addressed. The item groups were then compared with regard to the differences in response probabilities they created between the two classes. As expected, within each profile items of higher complexity were more difficult than those of low complexity. And while class 1 had a significantly higher probability of solving low complexity items (Mdn_{Class1} = 27.38,

$Mdn_{Class2} = 15.62$, $U = 97.000$, $p = .002$), the profiles did not differ significantly on items of high complexity ($Mdn_{Class1} = 12.64$, $Mdn_{Class2} = 10.36$), $U = 48.000$, $p = .49$.

Regarding the GPK sub-dimensions, only a cautious comparison is possible, as the number of items per dimension is too low to draw robust conclusions. Looking into the items' rankings within each class, one can see that both classes had a similar relative difficulty on the test items addressing assessment and lesson structuring. However, they differed in their ranking on adaptivity as well as on the classroom management and motivation items. While for class 1, adaptivity items were the easiest, they ranked only second for class 2, preceded by the classroom management and motivation items. The adaptivity items created not only qualitative differences of within class ranking, but also significant quantitative differences of probabilities between the classes ($Mdn_{Class1} = 13.11$, $Mdn_{Class2} = 5.89$), $U = 8.000$, $p = .003$.

4.6 Explaining profile membership

Since the item set PK26 created the largest quantitative and qualitative differences between the profiles, it was examined in more detail. The item set (Figure 10) required teachers to assign classroom exercises to the representation mode they illustrated according to Bruner's (1964) theory of representation. Bruner's theory of representation modes is considered part of teachers' GPK as it addresses general questions of knowledge creation and integration that are valid across subjects (Bruner 1966). However, in the literature the concept is commonly demonstrated based on mathematics education examples. Similarly, the item set PK26 was based on an exemplary context of teaching mathematics in primary education.

Figure 10. Task covering item group PK26

Imagine you are planning some exercises for a first grade class on the addition of numbers up to 20. Since the class is very heterogeneous with respect to their abilities, you want to offer three different types of exercises that correspond with Bruners' modes of representation: enactive, iconic, and symbolic.

Assign each of the following examples to one of the three exercise types.

Check one box in each row.

		enactive	iconic	symbolic	I do not know.
A.	Calculate: 8 + 12 = _____	☐1	☐2	☐3	☐4
B.	Arrange red and yellow gummy bears into mixed groups of the same size.	☐1	☐2	☐3	☐4
C.	On a worksheet, make circles around groups of 20 white and black birds.	☐1	☐2	☐3	☐4

One reason why some teachers might more likely have been assigned to the profile with a strength in adaptivity, was that mathematics teachers might have been more exposed to the topic of knowledge representation. As a result, they might have developed a specific strength that made it easier for them to solve the related items than for non-mathematics teachers. To test this hypothesis (RQ 3), a Pearson's Chi-Square and a Cramer's *V* test were conducted. The results displayed in Figure 10 strengthened our assumption, indicating a significant association between the teachers' subject-specific training background and their knowledge profiles (Chi-Square(2) = 32.0, $p < .001$, $n = 459$; Cramer's $V = .26$, $p < .001$). The odds-ratio underlined that if teachers had undergone at least basic training for mathematics teaching, their odds of belonging to class 1 was 4.13 times higher than if they had undergone other training.

Figure 11. Proportional distribution of teachers with and without training in mathematics teaching across the two latent classes
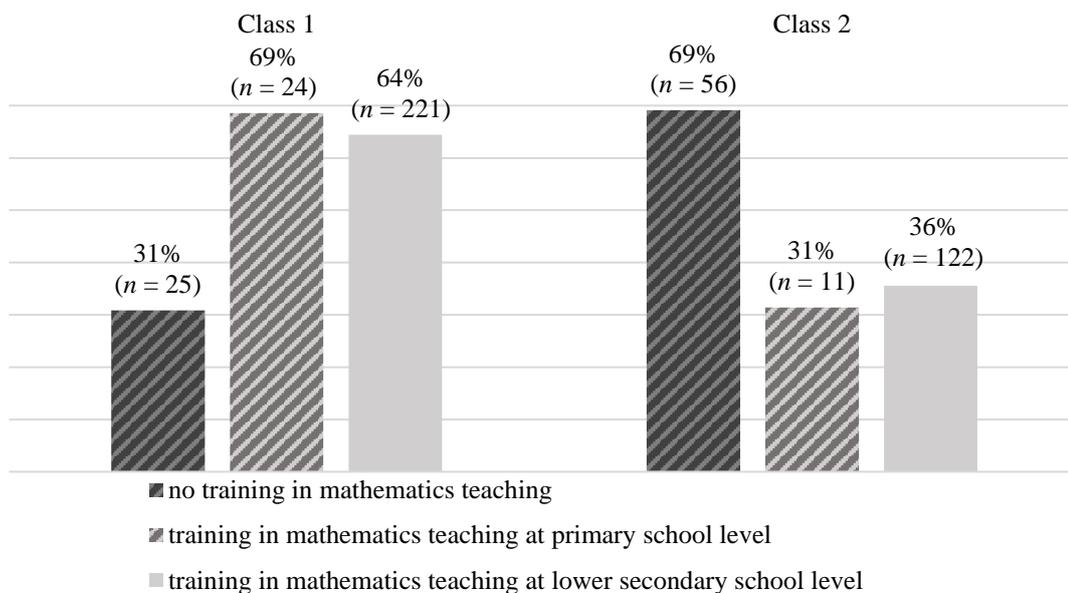
While the results hinted at a connection between the teachers' OTL and their probability of belonging to a certain profile, both the Cramer's *V* test and the distribution in Figure 10 underlined that the teachers' OTL could only partly explain class membership: some teachers were assigned to class 1 even though they had not undergone such training, whereas some of the teachers were assigned to class 2 despite their training in mathematics teaching.

4.7 Validating the GPK profiles

To validate the profiles (RQ 4), we tested whether teachers' of different GPK profiles showed different beliefs and instructional quality. Due to data collection constraints, these analyses

could only be applied to the sub-sample of mathematics teachers from TEDS-Validate, who were represented in both classes ($n_{\text{class1(TEDS-V)}} = 136; n_{\text{class2(TEDS-V)}} = 74$).

4.7.1 GPK profiles and teacher beliefs

As the classes did not only differ on the item set PK26, but also on other "adaptivity" items, we expect teachers to differ in their epistemological beliefs as well as their beliefs about teaching and learning. With a relative strength on items dealing with diversity in the classroom, teachers from class 1 could be expected to view the nature of mathematics less static and more dynamically. Additionally, they could be expected to show higher agreement to student-oriented, constructivist teaching views and lower agreement on teacher-oriented, transmissive views than those from class 2.

Results partly supported our hypotheses (see **Table *4.*** Mean values of agreement to belief scales by GPK profileTable 4): teachers from class 1 showed slightly higher dynamic and lower static beliefs than teachers from class 2. Differences in dynamic beliefs were only minor and non-significant. Differences in static beliefs, however, were significant and of small effect size, $t(208) = -2.793$, $p = .004$, Cohen' $d = -.38$. Class 1 further showed higher agreement with student-oriented and lower agreement with teacher-oriented views than class 2. Yet, only differences in teacher-oriented beliefs were significant, representing a small effect, $t(208) = -2.858$, $p = .005$, Cohen' $d = -.40$.

Table 4. Mean values of agreement to belief scales by GPK profile

Beliefs	Class 1 (n = 136)		Class 2 (n = 74)	
	M	*SD*	*M*	*SD*
Static	3.61**	0.75	3.90**	0.63
Dynamic	5.19	0.58	5.18	0.53
Teacher-oriented	2.44**	0.58	2.68**	0.57
Student-oriented	5.42	0.43	5.34	0.44

Note: Mean values marked with ** are significantly different between class 1 and 2, with $p < .01$.

4.7.2 GPK profiles and instructional quality

Based on the profiles' differences on "adaptivity" items, we expect teachers to show different levels of student support and cognitive activation. Both require dealing with learner diversity, even though at different points of the teaching and learning process (in problem induction vs. solving phases). Consequently, we expected class 1 to show a higher degree of student support and cognitive activation than class 2. Since classroom management is not specifically related

to adaptivity issues and the profiles hardly differed on classroom management items, we did not expect the classes to differ in the quality of their classroom management.

Table 5. Mean values of instructional quality scales by GPK profile

	Class 1 ($n = 51$)		Class 2 ($n = 23$)	
Instructional quality dimension	*M*	*SD*	*M*	*SD*
Classroom management	49.91	9.75	49.64	11.11
Student support	50.07	8.02	50.03	13.74
Cognitive activation	51.65*	10.38	46.30*	8.47

Note: Mean values marked with * are significantly different between class 1 and 2, with $p < .05$.

The results partly supported our hypotheses (Table 5). As expected, the two classes showed no significant differences in their classroom management. Unexpectedly, we found no significant differences between the two classes' level of student support either. Yet, in line with our hypothesis, class 1 showed a significantly higher level of cognitive activation than class 2 (Cohen's $d = .58$).

5. Discussion

5.1 Summary and discussion

Our study examined teachers' GPK profiles as part of their pedagogical competence. It aimed to unfold whether teachers were best represented by a single profile of common shape or by several profiles with varying strengths and weaknesses. The results confirmed our hypothesis, revealing the best fit for a 2-class MRM and showing that teachers not only differed in their overall GPK levels, but also in the composition of their strengths and weaknesses.

The study further examined the nature of the profiles and person- or item-related aspects that could have caused their differences. As expected, the items' required level of cognitive complexity was related to the items' difficulties within each of the profiles. Interestingly, despite their different overall GPK levels, the profiles had similar probabilities of solving highly complex items. However, they differed on items of low complexity, possibly indicating that other aspects or processes underlie these items and make them more difficult for some persons than others.

While many items had a similar probability of being solved, the profiles differed by an average of 35 percentage points on the test's adaptivity items. The profiles showed not only level, but also shape differences, as indicated by the items' difficulty ranking. Within class 1, these items were easiest, while, within class 2, items from the "classroom management and motivation"

dimension were easier. While such comparisons help describe the profiles' nature, they need to be treated with caution, as the number of items per dimension is not very high ($3 \leq n \leq 12$).

Three items related to Bruner's' (1964) theory of representation modes had a notable effect on the profiles, creating differences in both level and shape. The theory addresses instructional psychology relevant across teaching subjects. However, it is particularly regarded in the context of teaching mathematics and typically illustrated using mathematics examples, even in general pedagogy. Since the task in our test, too, was anchored in such a context, the next step of our analysis focused on teachers' education background and teaching subjects. It revealed that teachers who had undergone at least basic training in teaching mathematics were assigned significantly more often (yet not exclusively) to the profile that had almost no difficulty in solving the items.

The results are in line with other findings showing a link between OTL and GPK. They imply that mathematics teachers might be more likely to learn about representation modes as part of their teacher education and, thus, have no difficulty in solving corresponding tasks. One possible explanation could be that representation modes are not only addressed in OTL about general pedagogy but also in specific OTL about teaching methods in mathematics. The items might represent a good example of how closely linked certain GPK facets can be to MPCK.

Regardless of their training background, one third of the non-mathematics teachers was assigned to the higher performing profile and vice-versa. This hints at the possibility that additional aspects underlie the profile differences. One explanation might lie in the items' linguistic features: the item set is the only one written in a scientific language register. "Representation modes," "iconic," and "enactive" are not terms typically used in every-day situations. This would match findings by Klemenz and König (2019), according to which pre-service teachers had more difficulty solving items in a scientific language register than items using every-day or technical terms. However, this aspect could not be tested in our study, as the short version of the GPK test did not comprise enough items per language register. It would, thus, be valuable to reapply the MRM to the extended version of the instrument.

Despite the fact that the profiles' differences in overall GPK were rather small, the differences also showed in teachers' beliefs and instructional practice – thus, providing first evidence of the profiles' validity. A strength in adaptivity knowledge was associated with less static and teacher-oriented beliefs. At the same time, it was positively associated with cognitive activation levels during instruction. No significant differences were found in teachers' classroom management and student support levels – which was expected for the first, but not the latter. This might be explained by the fact that dealing with student diversity was only one aspect of

the student support scale, alongside socio- emotional support aspects. Thus, it might not have had sufficient weight in this dimension for the adaptivity knowledge to create a significant difference. In future studies, this could be addressed by implementing observational protocols covering cognitive student support in more detail.

5.2 Limitations

As mentioned previously, the brevity of the test version and the ratios of items per content dimension or item feature limits our analysis, requiring a certain degree of caution when interpreting the results. Another limitation of the study is the scarcity of background information on our sample. We had almost no additional information on the teachers without mathematics-specific training. Yet, their teaching subjects might help understand why some of them were assigned to the higher performing profile while others were not (e.g., if they taught subjects more closely related to mathematics such as natural sciences). Similarly, only little was available on the mathematics teachers' training background, so that we could not examine their OTL in more detail. Also, it was not possible to include a measure for CK or PCK, leaving open the question of the mediating role of subject-specific knowledge, e.g., between identified GPK profiles and cognitive activation or static beliefs.

5.3 Conclusion and outlook

Overall, our study demonstrates the potential of exploring profiles of teachers' GPK in their level and shape *simultaneously*. It illustrates how MRM can be used to identify classes that might otherwise be overlooked. Future research should address the limitations of this study by reapplying the MRM to the extended test version and a sample providing more information on teachers' personal background. A closer look could be taken at teachers' individual OTL patterns in order to relate them to specific characteristics of their competence profiles. Similarly, more extensive measures of GPK-related instructional quality dimensions could be used. With a specific focus on adaptivity, future research could take into account other adaptivity-related processes such as lesson planning to examine a more complex model that includes adaptivity-related knowledge and beliefs, their effect on teachers' situation-specific lesson planning and their instruction (see, as a first approach, König, Bremerich-Vos et al. under review).

Once the limitations are addressed, such findings could inform teacher training. Teacher education should ensure that pedagogical OTL provide examples for a diverse set of contexts, in order to prevent certain topics from being perceived as subject-specific and, potentially, as irrelevant for other subjects. More generally, such a profile analysis could provide (future)

teachers with recommendations on what to focus for their professional development. It could inform them on whether they need to increase their overall GPK levels or whether they should focus on certain contents or cognitive processes – and consequently allow for a more individualized professional development.

Acknowledgement

References

Alles, M., Seidel, T., & Gröschner, A. (2018). Establishing a positive learning atmosphere and conversation culture in the context of a video-based teacher learning community. *Professional Development in Education, 11*(1), 1–14.

Anderson, L., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Ayala, R. J. de, & Santiago, S. Y. (2017). An introduction to mixture item response theory models. *Journal of school psychology, 60,* 25–40.

Baghaei, P., & Carstensen, C. H. (2013). Fitting the mixed Rasch model to a reading comprehension test. Identifying reader types. *Practical Assessment, Research & Evaluation, 18*(5), 1–13.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 9*(4), 469–520.

Blömeke, S. (2012). Does greater teacher knowledge lead to student orientation? The relationship between teacher knowledge and teacher beliefs. In J. König (Ed.), *Teachers' pedagogical beliefs. Definition and operationalisation - connections to knowledge and performance - development and change* (pp. 15–35). Münster: Waxmann.

Blömeke, S., & Kaiser, G. (2012). Homogeneity or heterogeneity? Profiles of opportunities to learn in primary teacher education and their relationship to cultural context and outcomes. *ZDM Mathematics Education, 44*(3), 249–264.

Blömeke, S., Paine, L., Houang, R. T., Hsieh, F.-J., Schmidt, W. H., Tatto, M. T., et al. (2008). Future teachers' competence to plan a lesson. First results of a six-country study on the efficiency of teacher education. *ZDM Mathematics Education, 40*(5), 749–762.

Blömeke, S., Buchholtz, N., Suhl, U., & Kaiser, G. (2014). Resolving the chicken-or-egg causality dilemma: The longitudinal interplay of teacher knowledge and teacher beliefs. *Teaching and Teacher Education, 37,* 130–139.

Brühwiler, C. (2014). *Adaptive Lehrkompetenz und schulisches Lernen. Effekte handlungssteuernder Kognitionen von Lehrpersonen auf Unterrichtsprozesse und Lernergebnisse der Schülerinnen und Schüler* (Pädagogische Psychologie und Entwicklungspsychologie, vol. 91). Münster: Waxmann.

Brühwiler, C., Hollenstein, L., Affolter, B., Biedermann, H., & Oser, F. (2017). Welches Wissen ist unterrichtsrelevant? Validierung dreier Messinstrumente zur Erfassung des pädagogisch-psychologischen Wissens. *Zeitschrift für Bildungsforschung, 7*(3), 209–228.

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist, 19,* 1–15.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Depaepe, F., & König, J. (2018). General pedagogical knowledge, self-efficacy and instructional practice: Disentangling their relationship in pre-service teacher education. *Teaching and Teacher Education, 69,* 177–190.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development. Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199.

Dicke, T., Parker, P. D., Holzberger, D., Kunina-Habenicht, O., Kunter, M., & Leutner, D. (2015). Beginning teachers' efficacy and emotional exhaustion. Latent changes, reciprocity, and the influence of professional knowledge. *Contemporary Educational Psychology, 41,* 62–72.

Doyle, W. (1985). Recent research on classroom management. Implications for teacher preparation. *Journal of Teacher Education, 36*(3), 31–35.

Embretson, S. (2007). Mixed Rasch models for measurement in cognitive psychology. In M. Davier, & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models. Extensions and applications* (Statistics for Social and Behavioral Sciences, pp. 235–253). New York, NY: Springer Science + Business Media LLC.

Eye, A. von. (1999). Kognitive Komplexität - Messung und Validität. *Zeitschrift für Differentielle und Diagnostische Psychologie, 20*(2), 81–96.

Gräsel, C., & Trempler, K. (Eds.). (2017). *Entwicklung von Professionalität pädagogischen Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven*. Wiesbaden: Springer Fachmedien Wiesbaden.

Grothus, I., Renz, M., Rzejak, D., Schlamp, K., Daschner, P., Imschweiler, V., et al. (2018). Recherchen für eine Bestandsaufnahme der Lehrkräftefortbildung in Deutschland. Ergebnisse des Projektes Qualitätsentwicklung in der Lehrkräftefortbildung. *forum Lehrerfortbildung,* (47). Teil 1.

Hartig, J., & Frey, A. (2012). Konstruktvalidierung und Skalenbeschreibung in der Kompetenzdiagnostik durch die Vorhersage von Aufgabenschwierigkeiten. *Psychologische Rundschau, 63*(1), 43–49.

Jentsch, A., Schlesinger, L., Heinrichs, H., Kaiser, G., König, J., & Blömeke, S. (under review). Unterrichtsqualität unter einer mathematikdidaktischen Perspektive – Konzeptualisierung, Messung und Validierung. *Journal für Mathematik-Didaktik*.

Kleickmann, T., Hardy, I., Pollmeier, J., & Möller, K. (2011). Zur Struktur naturwissenschaftlichen Wissens von Grundschulkindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 43*(4), 200–212.

Klemenz, S., & König, J. (2019). Modellierung von Kompetenzniveaus im pädagogischen Wissen bei angehenden Lehrkräften. Zur kriterialen Beschreibung von Lernergebnissen der fächerübergreifenden Lehramtsausbildung. *Zeitschrift für Pädagogik, 65*(3).

König, J. (2014). *Designing an international instrument to assess teachers' general pedagogical knowledge (GPK): Review of studies, considerations, and recommendations*. Technical paper prepared for the OECD Innovative Teaching for Effective Learning (ITEL) - Phase II Project: A Survey to Profile the Pedagogical Knowledge in the Teaching Profession (ITEL Teacher Knowledge Survey), Paris: OECD.

König, J. (2015). Measuring classroom management expertise (CME) of teachers. A video-based assessment approach and statistical results. *Cogent Education, 2*(1), 1–15.

König, J., & Blömeke, S. (2010). *Pädagogisches Unterrichtswissen (PUW). Dokumentation der Kurzfassung des TEDS-M-Testinstruments zur Kompetenzmessung in der ersten Phase der Lehrerausbildung,* Berlin: Humboldt-Universität.

König, J., & Kramer, C. (2016). Teacher professional knowledge and classroom management. On the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM Mathematics Education, 48*(1-2), 139–151.

König, J., & Pflanzl, B. (2016). Is teacher knowledge associated with performance? On the relationship between teachers' general pedagogical knowledge and instructional quality. *European Journal of Teacher Education, 39*(4), 419–436.

König, J., & Seifert, A. (Eds.). (2012). *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung*. Münster u.a.: Waxmann.

König, J., Bremerich-Vos, A., Buchholtz, C., & Glutsch, N. (under review). General pedagogical knowledge, pedagogical adaptivity in written lesson plans, and self-reported instructional practice: On the modelling of planning competence among preservice teachers. *Journal of Teacher Education*.

König, J., Kaiser, G., Blömeke, S., Jentsch, A., Schlesinger, L., Nehls, C., et al. (under review). Teaching and learning in the lower secondary mathematics classroom. Analyzing the links between pedagogical competence, instructional quality, and student achievement.

König, J., Blömeke, S., Paine, L., Schmidt, W. H., & Hsieh, F.-J. (2011). General pedagogical knowledge of future middle school teachers. On the complex ecology of teacher education in the United States, Germany, and Taiwan. *Journal of Teacher Education, 62*(2), 188–201.

König, J., Blömeke, S., & Kaiser, G. (2015). 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, 13*(2), 331–350.

Kunter, M., & Klusmann, U. (2010). Die Suche nach dem kompetenten Lehrer - ein personenzentrierter Ansatz. In W. Bos (Ed.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (pp. 207–230). Münster: Waxmann.

Lauermann, F., & König, J. (2016). Teachers' professional competence and wellbeing. Understanding the links between general pedagogical knowledge, self-efficacy and burnout. *Learning and Instruction, 45,* 9–19.

Lenske, G., Wagner, W., Wirth, J., Thillmann, H., Cauet, E., Liepertz, S., et al. (2016). Die Bedeutung des pädagogisch-psychologischen Wissens für die Qualität der Klassenführung und den Lernzuwachs der Schüler/innen im Physikunterricht. *Zeitschrift für Erziehungswissenschaft, 19*(1), 211–233.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767–778.

Lohse-Bossenz, H., Kunina-Habenicht, O., Dicke, T., Leutner, D., & Kunter, M. (2015). Teachers' knowledge about psychology. Development and validation of a test measuring theoretical foundations for teaching and its relation to instructional behavior. *Studies in Educational Evaluation, 44,* 36–49.

Lohse-Bossenz, H., Holzberger, D., Kunina-Habenicht, O., Seidel, T., & Kunter, M. (2018). Wie fach(un)abhängig ist bildungswissenschaftliches Wissen? Messinvarianz und fachspezifische Unterschiede. *Zeitschrift für Erziehungswissenschaft, 21*(5), 991-1019.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological methods, 10*(1), 21–39.

Lubke, G. H., & Neale, M. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes. Class invariance of parameters of factor mixture models. *Multivariate behavioral research, 43*(4), 592–620.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis, 17*(3), 305–322.

Ministerium für Schule, Jugend und Kinder des Landes Nordrhein-Westfalen. (2004, 30 June). *Didaktisches Grundlagenstudium Mathematik. für das Lehramt an Grund-, Haupt- und Realschulen sowie den entsprechenden Jahrgangsstufen der Gesamtschulen.* Fächerspezifische Vorgaben.

Minnameier, G., Hermkes, R., & Mach, H. (2015). Kognitive Aktivierung und Konstruktive Unterstützung als Prozessqualitäten des Lehrens und Lernens. *Zeitschrift für Pädagogik, 61*(6), 837–856.

Morin, A. J. S., & Marsh, H. W. (2014). Disentangling shape from level effects in person-centered analyses. An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(1), 39–59.

Muthén, L., & Muthén, B. (1998-2012). *Mplus. Statistical analysis with latent variables.* User's guide. https://www.statmodel.com/download/usersguide/ Mplus%20user%20guide%20Ver_7_r3_web.pdf.

Nehls, C., König, J., Kaiser, G., Klemenz, S., Ross, N., & Blömeke, S. (under review). Validierung des TEDS-M-Instruments zum pädagogischen Wissen bei Validierung des TEDS-M-Instruments zum pädagogischen Wissen bei berufserfahrenen Mathematiklehrkräften. *Diagnostica.*

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in Latent Class Analysis and Growth Mixture Modeling. A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535–569.

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality. The German framework of Three Basic Dimensions. *ZDM Mathematics Education, 47,* 97.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*(1), 103–124.

Rost, J. (1990). Rasch models in latent classes. An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*(3), 271–282.

Rost, J. (2007). Definition von Kompetenzniveaus mit Hilfe von Mischverteilungsmodellen. In B. Beck, & E. Klieme (Eds.), *Sprachliche Kompetenzen. Konzepte und Messung: DESI-Studie (Deutsch Englisch Schülerleistungen International)* (Beltz-Pädagogik, pp. 100–106). Weinheim: Beltz Verlag.

Schulze-Stocker, F., Holzberger, D., Kunina-Habenicht, O., Terhart, E., & Kunter, M. (2016). Spielen Studienschwerpunkte wirklich eine Rolle? Zum Zusammenhang von bildungswissenschaftlichen Studienschwerpunkten, selbst eingeschätzten Kenntnissen und gemessenem Wissen am Ende eines Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft, 19*(3), 599–623.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. (Beschluss der Kultusministerkonferenz vom 16.12.2004),* Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK).

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2017). *Sachstand in der Lehrerbildung (Stand: 07.03.2017).*

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–23.

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework*. East Lansing, MI: Teacher Education and Development International Study Center.

Voss, T., & Kunter, M. (2013). Teachers' general pedagogical/psychological knowledge. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (Mathematics teacher education, vol. 8, pp. 207–228). New York: Springer.

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.

Voss, T., Kunter, M., Seiz, J., Hoehne, V., & Baumert, J. (2014). Die Bedeutung des pädagogisch-psychologischen Wissens von angehenden Lehrkräften für die Unterrichtsqualität. *Zeitschrift für Pädagogik, 60*(2), 184–201.

Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften. Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft, 18*(2), 187–223.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Ashland, U.S.: Hogrefe & Huber Publishers.

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* (Cassel education, pp. 104–124). London: Cassell.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues and Answers Report REL 2007–No. 033), Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.