Teaching and Teacher Education 56 (2016) 35–46

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

The relation between content-specific and general teacher knowledge and skills

Sigrid Blömeke [a,*], Andreas Busse [b], Gabriele Kaiser [b], Johannes König [c], Ute Suhl [d]

[a] University of Oslo/Centre for Educational Measurement (CEMO), N – 0318 Oslo, Norway
[b] University of Hamburg, D – Hamburg, Germany
[c] University of Cologne, D – Köln, Germany
[d] Humboldt University of Berlin, D – 10099 Berlin, Germany

HIGHLIGHTS

- 171 teachers were tested on content-related and general knowledge and skills.
- CFA distinguished between levels of generalizability across teaching situations.
- Domain-specific, assessment-specific and one-dimensional cognitive models fit worse.
- Grades in the teaching exam were positively related to situation-specific skills.
- General cognitive abilities were positively related to knowledge and skills.

ARTICLE INFO

Article history:
Received 30 June 2015
Received in revised form
31 January 2016
Accepted 7 February 2016
Available online 18 February 2016

Keywords:
Content knowledge
Pedagogical content knowledge
General pedagogical knowledge
Situation-specific skills
Classroom management
Mathematics education

ABSTRACT

The relation between teacher knowledge and skills and how these were influenced by teacher education was examined with 171 secondary mathematics teachers. Six paper-and-pencil and video tests were applied to assess content knowledge, pedagogical content knowledge and general pedagogical knowledge as well as diagnostic, teaching and classroom management skills. It was hypothesized that the relation between these six cognitive facets was best approximated by distinguishing between levels of generalizability across different mathematics teaching situations. The data strongly supported this model in confirmatory factor analyses. The data also revealed the hypothesized differential relations between teacher cognitions and teacher education.

1. Introduction

In-depth research on teacher cognition that includes a broad range of knowledge and skills facets and does not only focus on, for example, content knowledge (e.g., Baumert et al., 2010; Hill, Rowan & Ball, 2005) or classroom-management skills (Evertson & Weinstein, 2009) is still scarce. Most studies available focused either on content-specific *or* on general pedagogical facets, and within these again either on knowledge *or* on skills. This research gap exists although it is well known that teacher performance in the classroom is based on the integration of a range of cognitive resources in addition to beliefs, values and motivation (Schoenfeld, 2010). How precisely the different facets of teachers' knowledge and skills are related to each other is therefore not known because they have rarely been assessed in one study.

Furthermore, due to the challenges related to direct testing of teachers, self-reported data is still the most common approach in teacher research although their reliability flaws are widely known. A first aim of the present study was against this background to directly test different facets of teacher cognition and then using this data to clarify their relation to each other.

A second aim of the study was to examine effects of different types of teacher education on the cognitive structure identified. In Germany where the present study took place, mathematics teachers

* Corresponding author.
E-mail address: sigribl@cemo.uio.no (S. Blömeke).

http://dx.doi.org/10.1016/j.tate.2016.02.003

achieve their credentials through two different pathways. Those with preparation as “upper- and lower-secondary teachers” (includes typically teaching in grades 5 through 12) tend to have stronger mathematical preparation in high school, spend more time studying advanced mathematics in college and have more experience in mathematics classrooms as part of their preparation compared to those prepared as “lower-secondary teachers” only (includes typically teaching in grades 5 through 9). Such differences in training may result in different relationships between knowledge and skills. In distinguishing between different groups of teachers, the study will therefore not only provide insight into the structure of teacher cognition but also into potential effects of teacher education. The more information about the structure of teacher cognition is available and how it is related to teacher education, the better initial teacher education and professional development activities can be developed tailored to teachers' needs.

2. Conceptual framework

Blömeke, Gustafsson and Shavelson (2015) integrated research on teacher expertise into teacher knowledge frameworks and distinguished between teacher knowledge as rather stable cognitive resources generalizable across different mathematics teachingsituations on the one hand and cognitive skills which are more related to very specific classroom situations and, thus, more variable on the other hand. This integrated framework served as a point of reference for the present study.

2.1. Facets of teacher knowledge

Research on teacher knowledge and how it is structured has become an important research field during the past 10 years, in particular with respect to mathematics teachers (e.g., Hill, Rowan, & Ball, 2005). Inspired by Shulman's (1987) conceptualization, Schoenfeld and Kilpatrick (2008) developed a framework that distinguished between two facets of content-specific knowledge, namely *mathematics content knowledge* (MCK) and *mathematics pedagogical content knowledge* (MPCK). MCK includes fundamental mathematical definitions, concepts, algorithms, and procedures whereas MPCK includes knowledge about how to teach these mathematical concepts and procedures to students.

One of the earliest large-scale teacher assessments – the international “Teacher Education and Development Study: Learning to Teach Mathematics” (TEDS-M) – used this framework (Tatto et al., 2008) and assessed MCK and MPCK of secondary mathematics teachers from 16 countries directly with paper-and-pencil tests using mainly multiple-choice items. MCK covered from a higher level the mathematical content of the grades the teachers would teach. MPCK covered the conveyance of mathematical concepts and methods. On the basis of Anderson and Krathwohl's (2001) framework of cognitive processes, TEDS-M items assessed knowing and remembering MCK and MPCK as well as understanding and applying MCK and MPCK (Döhrmann, Kaiser, & Blömeke, 2012). Higher-order cognitive processes such as creating and generating MCK and MPCK strategies were only rarely covered by the TEDS-M items. Döhrmann et al. (2012) characterized the tests therefore as one that assessed predominantly declarative knowledge. Results from TEDS-M pointed to medium or strong relations between MCK and MPCK in all countries but one (Pearson's $r = 0.37$–0.70; Blömeke, Kaiser & Lehmann, 2010).

Shulman (1987) had also conceptualized a general facet of teacher knowledge, namely *general pedagogical knowledge* (GPK), defined as “broad principles and strategies for classroom management and organization that transcend subject matter” (p. 8). Germany, Taiwan and the U.S. developed a corresponding paper-and-pencil test with mostly open-ended items and brief written classroom scenarios in the context of TEDS-M (Blömeke, Kaiser & Lehmann, 2010). The test covered the same cognitive processes as the main study but included also items and scenarios covering Anderson and Krathwohl's (2001) highest level of creating and generating instructional strategies. Blömeke, Kaiser & Lehmann (2010) characterized the test therefore as one that assessed declarative but also procedural knowledge. Results pointed to low to medium relations between GPK and MPCK or MCK ($r = 0.14$–0.30 or 0.11–0.29).

2.2. Facets of cognitive skills

The body of studies examining teachers' cognitive skills has recently grown, in particular their skills to *perceive*, *interpret* and to make *decisions* with respect to general *classroom management* (*P-I-D CM*; Gold, Förster, & Holodynski, 2013; Stürmer, Könings, & Seidel, 2012) but also with respect to teachers' skills to *perceive*, *interpret* and to make *decisions* about *mathematics instruction* (*P-I-D math*). Mathematics instruction is to our knowledge the only content domain where this has been examined systematically. Kersting (2008) showed that mathematics teachers' skills of perceiving mathematics instruction were significantly positively correlated with their MCK. Sherin, Jacobs, and Philipp (2011) showed that perceiving mathematics instruction in turn predicted these teachers' performance in the classroom (see also Star & Strickland, 2008).

Drawing on expertise research, Krauss and Brunner (2011) distinguished another skill facet from those pointed out above, and this was mathematics teachers' *diagnostic* skills to identify student errors in *mathematics* (MDiagnose). Their data revealed that mastery of this skill facet could be identified through a speed component because the time teachers needed to diagnose students' mathematical errors differed significantly. This result reflects that experts can make rapid judgments based on a rich knowledge and skills base because they dispose of more cognitive chunks than novices (Clark & Lampert, 1986). Mentally grouping classroom situations overcomes the limits of short-term memory so that, with extended experience, experts can retrieve information more quickly and from a broader repertoire of critical incidents than novices. Expert teachers have therefore an idea about problems already prior to a lesson, for example about typical student errors and on which parts of a student solution to focus (Bromme, 1992). This skill facet is therefore more independent from specific classroom situation and can be regarded as rather generalizable. According to Krauss and Brunner (2011), this skill is significantly related to MCK.

2.3. Research gaps with respect to teacher knowledge and skills

Although a number of studies on the different facets of teacher knowledge or skills exist, it is widely unknown how general pedagogical and content-specific skills are related to each other and how these in turn are related to general pedagogical and content-specific knowledge facets because the facets are rarely examined within one study. The gap between general and content-specific research may partly go back to expertise research itself. Already De Groot (1946/1978) had defined expertise as a domain-specific construct. Later research confirmed that experts have difficulties to transfer their knowledge and skills from one domain to another (Glaser & Chi, 1988; Van Overschelde, Rawson, Dunlosky, & Hunt, 2005). However, the demands teachers are confronted with in a classroom require to combine a content-specific perspective on learning and instruction with a general pedagogical perspective on classroom management (Fauth, Decristan, Rieser, Klieme, & Büttner, 2014). The definition of a “domain” may therefore be different in the case of teachers than else in expertise research.

Direct testing of teacher cognition represents a challenge though. To our knowledge, all studies that assessed teacher *skills* did this with video- or computer-based assessments. This means that bias may have been introduced by applying the same test format instead of changing it by, for example, using also paper-and-pencil tests. Such a methods bias is a well-known problem in educational and psychological research (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Using the same format may introduce artificial covariance or – vice versa – using different formats may introduce artificial non-covariance. The present study combined therefore different assessment formats to examine constructs on the same level of generalizability (either situation-specific or generalizable across mathematics teaching situations) and can therefore test whether the risk of methods bias exists.

2.4. Hypotheses about the structure of teacher cognition

A model that hypothesized the structure of teacher cognition based on the generalizability of each facet across classroom situations (relatively stable vs. situation-specific) was tested against two models that hypothesized competing structures (general pedagogical vs. content-specific facets, paper-pencil vs. computer-based assessments).

2.4.1. Distinction of teachers' cognitive facets according to level of generalizability (H1)

We hypothesize that a model of teacher cognition fits best to the data that distinguishes between rather stable cognitive resources on the one hand and resources more related to specific classroom situations on the other hand (H1a; see Fig. 1). Both dimensions can be regarded as necessary preconditions of classroom performance and should correlate significantly positive (H1b) because they belong to the same overall construct.

MCK and MPCK were in the context of TEDS-M purposefully operationalized as stable facets of mathematics teachers' cognition and generalizable across different situations in a mathematics classroom. Whether a teacher is able to diagnose typical mathematical errors under time pressure (MDiagnose) can also be hypothesized to be rather independent of the specific classroom situation (Krauss & Brunner, 2011). We hypothesize therefore that MCK, MPCK and MDiagnose constitute the first dimension of teachers' cognition (H1c).

P-I-D CM and P-I-D math were by definition conceptualized to be of situated nature by covering perceptual, interpretation and decision-making skills with respect to specific classroom management or mathematics instruction situations. Similarly, the TEDS-M GPK assessment was purposefully developed to cover not only declarative but also procedural knowledge. We hypothesize therefore that these three indicators constitute the second dimension of teacher cognition (H1d).

2.4.2. Competing hypotheses (H2–H4)

Considering that mathematics as the content taught and general pedagogical classroom demands could be regarded as distinct domains, a competing model could be hypothesized in which one dimension consists of the four mathematics-specific facets MCK, MPCK, MDiagnose and P-I-D math whereas the second dimension consists of the two general pedagogical indicators GPK and P-I-D CM (H2a; see Fig. 2).

Given that methodological artifacts represent a serious problem in educational and psychological research, another two-dimensional model could be hypothesized (H3) that represents the different assessment approaches used because such different assessment approaches may tap different constructs. MCK, MPCK and GPK were assessed with (digitalized)paper-and-pencil tests and could therefore build one dimension whereas P-I-D math, P-I-D CM and MDiagnose were assessed in a video-based way and could therefore build a second dimension (see Fig. 3).

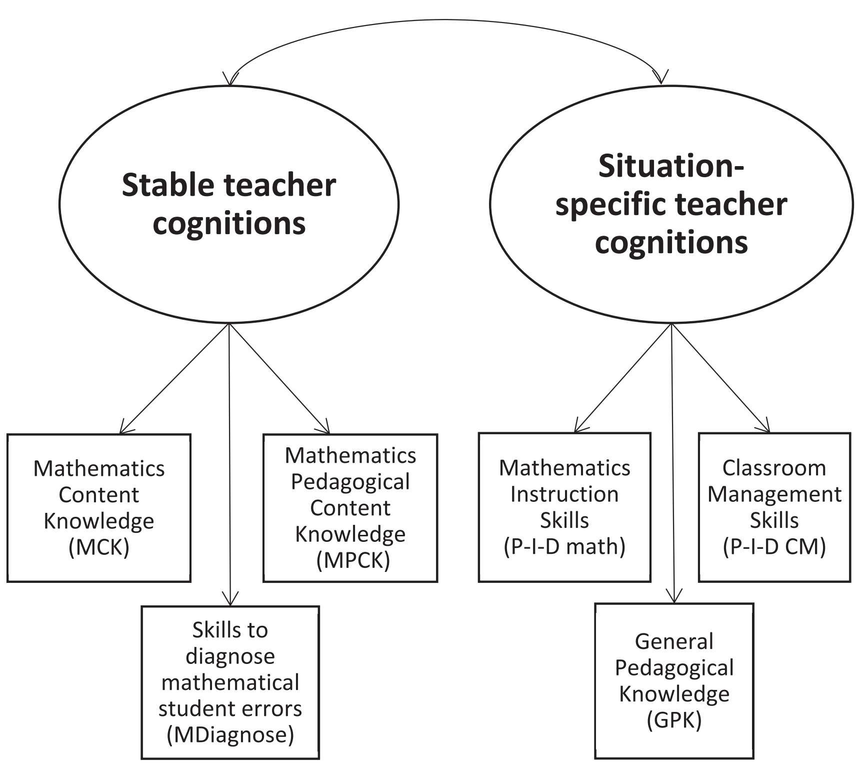


Fig. 1. Two-dimensional model of teachers' knowledge and skills distinguishing between levels of generalizability (H1).

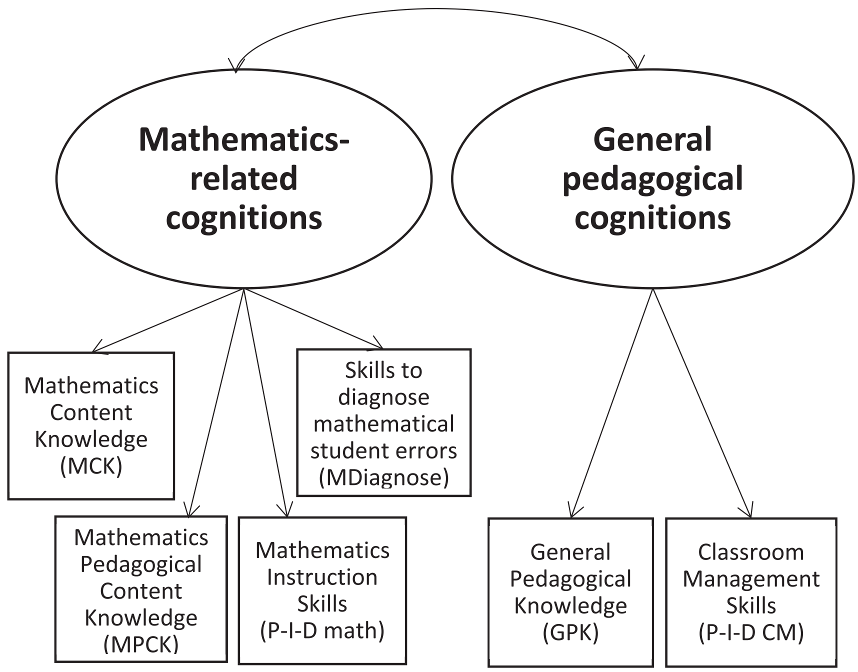


Fig. 2. Two-dimensional model of teacher cognitions distinguishing between domains (H2).

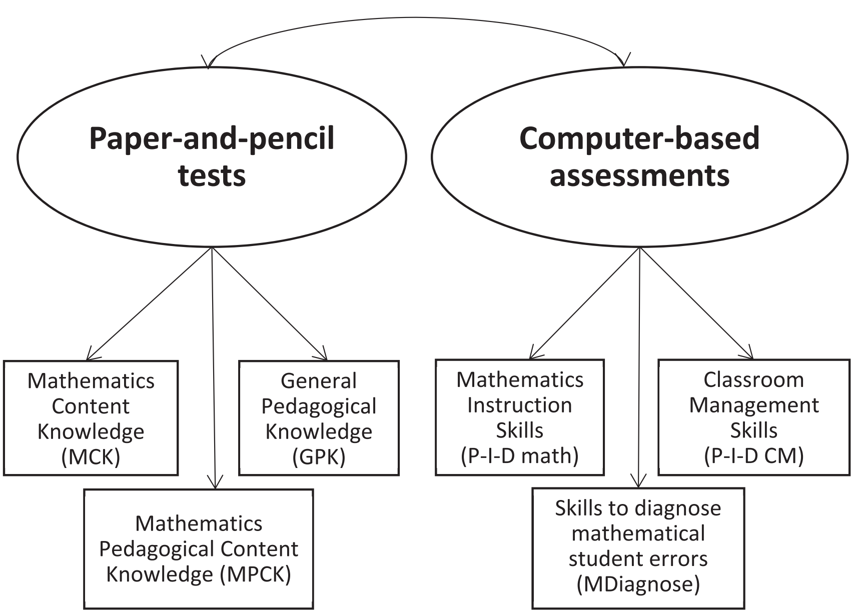


Fig. 3. Two-dimensional model of teacher cognitions distinguishing between assessment formats (H3).

Finally, it might be that a one-dimensional model is sufficient to describe teachers' cognitive facets. Some studies revealed that with increasing expertise the interrelation between cognitive facets becomes stronger so that different cognitive facets involved in solving a task could not be separated from each other anymore (Smith & Strahan, 2004). Such teachers would often excel on many indicators (Sabers, Cushing, & Berliner, 1991). Although this applies typically to more experienced teachers, which are not part of the study (see below), a one-dimensional model with the six facets MCK, MPCK, MDiagnose, GPK, P-I-D math and P-I-D CM should be tested to exclude this option (H4; see Fig. 4).

2.5. Two different types of teacher education

As mentioned, secondary teachers in Germany are trained as teachers for lower-secondary schools only or as teachers for lower- and upper-secondary schools. These schools represent different tracks (*Hauptschule*, *Realschule*, *Mittelschule* and integrated schools on the one hand and the *Gymnasium* on the other hand) and students are selected into these based on their achievement at the end of primary school.

The two groups of teachers undergo different types of teacher education (Blömeke, Kaiser & Lehmann, 2010). Whereas the practical training is roughly the same, university programs of lower-

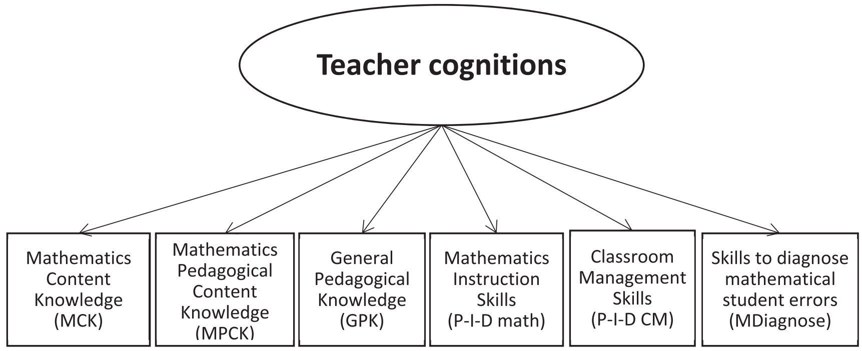


Fig. 4. One-dimensional model depicting teacher cognitions as homogenous (H4).

secondary school teachers have up to now in almost all German federal states been shorter than those of teachers expected to teach at upper-secondary levels[1] as well. Typically, the university programs for the first group last for 3.5 years whereas those for the latter last for 5 years (ibid.). In addition, programs for preparing lower-secondary teachers cover typically more subjects than programs preparing teachers for upper- and lower-secondary school. A program for the first group, may include for example three subjects besides a large share of general pedagogy instead of one major and one minor besides a small share of general pedagogy in case of the second group.

These characteristics result in fewer opportunities for prospective lower-secondary mathematics teachers to build up a stable mathematics-related cognitive base. Furthermore, self-selection of prospective teachers into these two programs differs substantially (ibid.). Lower-secondary teachers had on average less strong education in mathematics during high-school and a lower general grade-point average in the high-school exit exam.

All teachers have to take a practical exam at the end of their programs ("Zweites Staatsexamen", second state exam). It consists of a series of lesson observations by multiple raters, namely teacher educators and principals. The grade point average on this practical teaching exam provides, thus, information about teachers' actual classroom behavior.

2.6. Hypotheses about the relation of teacher education to teachers' cognitive dimensions (H5)

The differences in prerequisites for and the teacher education programs themselves should result in subgroup differences in terms of different levels of expertise. We hypothesize lower means of the lower-secondary teachers on the stable cognitive dimension compared to teachers prepared for teaching on the upper- and lower-secondary levels (H5a).

Furthermore, expertise research suggests that experts profit from a well-connected cognitive base (Sabers et al., 1991) which means that the relation between the two cognitive dimensions should be stronger in case of mathematics teachers for lower- and upper-secondary schools compared to teachers for lower-secondary schools only (H5b).

Practical training during teacher education can be expected to influence more strongly situation-specific skills than the knowledge base. It is therefore hypothesized that the grade point average in the final state exam is more strongly related to the situation-specific than to the stable cognitive dimension (H5c). Such a result would also confirm the practical significance of cognitive skills as assessed with our instruments for instructional processes.

3. Method

3.1. Participants

The sample consisted of 171 lower-secondary mathematics teachers from all 16 federal states in Germany who had undergone an initial teacher education of about 3.5–5 years depending on the program and the state where it took place, a practical training of about 1.5 years (called induction in some countries), and were now roughly in their third year in the profession. Initial teacher education in Germany is a typical university-based study program in general pedagogy and two or more subjects depending on the program and the state. The following practical training is characterized by elements described as deliberate practice by Ericsson (2005), namely, striving for improvement, receiving high-quality feedback by master teachers, and opportunities to repeat their performance.

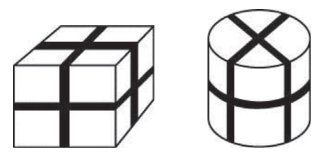
72 teachers were licensed to teach at lower-secondary schools only, 91 were licensed also to teach at upper-secondary schools. Overall, the mathematics teachers had about 8–10 years of university training, deliberate practice during induction, and practical experience. In 2008, the teachers had participated in the German TEDS-M study while they were in their final year of teacher education. After passing the exit exam, they became teachers. 62% of the randomly drawn TEDS-M sample had agreed to participate in further studies, about half of these could be found again four years later in 2012. They were contacted via email, phone calls or letters for the present study. The 171 mathematics teachers who finally agreed to participate in this study received an honorarium to compensate for their efforts.

Table 1 reports socio-demographic and educational characteristics of the sample. The teachers were on average 32 years old, almost 60% were female. About three quarters had attended an advanced mathematics class in high school. They finished high school on average with a good exit exam (grade point average of 2.1 on a scale from 1 as the best to 6 as the worst grade; 2.1 is also significantly above the national average; KMK, 2006). The first and second state exams, which took place at the end of initial teacher education and after their practical training was finished with good grades, too (grade point averages of 1.9 or 2.1, respectively). The teachers were from families with high cultural capital, more than half of them reported three or more shelves with books at home.

[1] This situation is currently changing but the change does not affect the relevance of the study because the change has not led to extended opportunities to acquire content knowledge in the formerly shorter programs but to extended school practice. The differences in qualification of the two teacher groups will therefore be even more pronounced in the future.

Two gift boxes wrapped with ribbon are shown below. Box A is a cube of side-length 10 cm.

Box B is a cylinder with height and diameter 10 cm each.



A B

Which box needs the longer ribbon? _______

Explain how you arrived at your answer

Code	Response	Item MFC703
	Correct Response	
20	Box A with a correct and complete explanation involving calculations of ribbon lengths *Example: Box A needs* $6 \times 20 = 120$ *cm ribbon. Box B needs* $4 \times 20 = 80$ *cm plus the circumference which is* 10π *.* $10\pi < 40$ *so Box A needs more ribbon.*	
21	Box A based upon a complete argument (with or without calculation) comparing the square and circumference (both of equal “width”) together **with a statement that the other lengths of ribbon are equal.** *Example: Box A because the circumference of a circle with diameter 10 is less than the perimeter of a square with side 10 and the other lengths of ribbon are the same.*	
	Partially Correct Response	
10	Box A with a correct and complete explanation as in Code 20 but with **one identifiable calculation error** (or use of a wrong formula) logically leading to Box A. *Example: Box A because Box A needs 120 cm and Box B needs* **60** + $10\pi < 120$.	
11	Box B with a correct and complete explanation as in Code 20 but with **one identifiable calculation error** (or use of a wrong formula) logically leading to Box B. *Example:* $80 + 10\pi = 120.4$ (rather than 111.4) > 120.	
12	Box A with an explanation that correctly calculates and compares the lengths of ribbon on each box that are **different** but **fails to mention that the other lengths of ribbon are the same.** *Example: Box A needs more ribbon because the circumference of the cylinder is* 10π *, which is less than the perimeter of the square,* 40.	
13	Box A with an explanation that correctly supports the choice of Box A but that is **limited** and/or **lacking detail**. *Example: Box A because Box B can fit inside Box A.*	
	Incorrect Response	
70	Box A **without** any explanation or calculation. *Example: Box A*	
71	Box A or B with an explanation based on a **conceptual error**. *Example: Box A but with an explanation based upon surface* ***area or volume****.*	
72	Box A or B with an explanation based on **incorrect and/or incomplete** ribbon lengths for **both** boxes. *Example: Box B because Box A needs* 60 *cm but box B needs more than* 80.	
73	Neither. The length of ribbon needed is the **same.**	
79	Other incorrect (including crossed out, erased, stray marks, illegible, or off task)	
	Non-response	
99	Blank	

Fig. 5. Example of a constructed-response MCK item.

The following problems appear in a mathematics textbook for <lower secondary school>.

1. [Peter], [David], and [James] play a game with marbles. They have 198 marbles altogether. [Peter] has 6 times as many marbles as [David], and [James] has 2 times as many marbles as [David]. How many marbles does each boy have?
2. Three children [Wendy], [Joyce] and [Gabriela] have 198 zeds altogether. [Wendy] has 6 times as much money as [Joyce], and 3 times as much as [Gabriela]. How many zeds does each child have?

(a) Solve each problem.
(b) Typically Problem 2 is more difficult than Problem 1 for <lower secondary> students. Give one reason that might account for the difference in difficulty level.

Fig. 6. Example of a constructed-response MPCK item (Part b).

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?*

Fig. 7. Example of an open-ended GPK item with an excerpt of an original response.

Table 1
Descriptive characteristics of the sample.

Age (M, SD)	32.1 (5.9)
Percentage of female teachers	59%
Percentage of teachers with 3 or more bookshelves at home	53%
Percentage with advanced mathematics in upper-secondary school	73.3%
Grade in high school exit exam (M, SD)	2.1 (0.6)
Grade in first state exam (M, SD)	1.9 (0.5)
Grade in second state exam (M, SD)	2.1 (0.7)
Percentage of "upper- and lower secondary teachers"	55.8%

Note. M = mean, SD = standard deviation. All three grade scales range from 1 as the best grade to 6 as the worst with 4 (just passed) as the threshold between pass and fail.

About half of the teachers worked at schools with lower- and upper secondary classes (*Gymnasium*), whereas the other half worked at lower-secondary schools (*Hauptschule*, *Realschule*, *Mittelschule* or integrated schools).

3.2. Instruments

3.2.1. MCK

MCK was assessed with an abbreviated version of the formerly rather long TEDS-M test (Tatto et al., 2012). To lessen the test burden, given that the mathematics teachers were supposed to work on six assessments, the initial 76 items were reduced to 30 items without loss of reliability (see Table 2 for properties of all assessments). The items covered number, algebra, and geometry as core areas of school mathematics and with a few items also data. All TEDS-M paper-and-pencil tests were provided in a digitalized way to facilitate test administration.

Fig. 5 presents one example item together with a coding scheme excerpt. The item requires the comparison of properties of three-dimensional objects. Test-takers need to identify whether the ribbon of the cube or the cylinder is longer and to explain the choice adequately. Answers were regarded as correct and complete if box A was identified and the choice was explained by correctly calculating both ribbon lengths and the length of the circumference or by comparing the circumferences of the circle and the square regarding that the remaining lengths are equal for both boxes. Incomplete answers as well as answers with minor errors were valued as partial solutions. Answers were considered wrong if misconceptions became apparent (for example calculating the wrong parameter such as area or volume), if the lengths of both ribbons were calculated incorrectly, if teachers expected the same length for both ribbons or if a correct answer did not include an explanation because this was the main focus for this task.

3.2.2. MPCK

MPCK was assessed with a digitalized version of the original TEDS-M assessment (Tatto et al., 2012). The items covered aspects of curricular and planning knowledge as well as knowledge about how to teach mathematics. These two sets of items were given approximately equal weight. Part b of Fig. 6 presents one example item that covers knowledge how to teach algebra. At a first glance, the two problems look very similar because each time the overall number and the relation of objects each child possesses are known and it has to be calculated how many objects each child possesses. However, in the first problem the relational information refers in both cases to David so that the *number* of Peter's and James' possessions can directly be calculated whereas in the second problem the *relation* of Joyce's and Gabriela's number of possessions has to be calculated first. Teachers need to be aware that the latter problem is more difficult for students than the first one.

Table 2
Overview about instrument properties.

Construct	Method	No. of items	Item format	Reliability
MCK	Digitalized paper-and-pencil test	30	MC, CR	0.73
MPCK	Digitalized paper-and-pencil test	29	MC, CR	0.76
GPK	Digitalized paper-and-pencil test	39	MC, CR, OE	0.75
P-I-D CM	Digitalized video assessment	40	RS, OE	0.70
P-I-D math	Digitalized video assessment	34	RS, OE	0.67
MDiagnose	Computer-based	16	MC	0.73

Note. MC = Multiple-choice items, CR = constructed-response items, OE = open-ended items, RS = rating scales. The six scores were scaled separately from each other.

3.2.3. GPK

GPK was assessed with an abbreviated and digitalized version of the instrument developed in the context of TEDS-M by Germany, Taiwan and the U.S. (König & Blömeke, 2009). The 39 items used of the initial 77 items were fairly equally distributed across teacher tasks such as lesson planning, dealing with heterogeneity, motivation, classroom management, and assessment. Fig. 7 presents one example item from this test. The teachers were asked to support a future teacher and evaluate her lesson. This is a typical challenge during a peer-led teacher education practicum, but practicing teachers are also regularly required to analyze and reflect on their own as well as their colleagues' lessons. The item measured knowledge about lesson planning. The predominant cognitive process was to generate fruitful questions related to instruction. A question was accepted as correct if it addressed the context (e.g., prior knowledge of students), the input (e.g., learning objectives), the process (e.g., teaching methods used), or the output of the lesson (e.g., student achievement).

All tests had been validated as part of TEDS-M (Tatto et al., 2008, 2012). The item development had to follow the study's conceptual framework to ensure content validity. To avoid cultural bias, items had to be sent in from all participating countries. The item pool was reviewed by large groups of experts, on both an international level and within the participating countries. Translation processes had to follow strict rules and all national research coordinators had to approve the final versions of the instruments. High psychometric quality was ensured, including evidence of internal consistency, score reliability, and measurement invariance (Tatto et al., 2012). Exploratory and confirmatory factor analyses were carried out to assess the fit of each scale to the data with reference to the conceptual framework. The full set of released TEDS-M items is available at tedsm@msu.edu. Documentation of parameter estimates of these items is available as well (Laschke & Blömeke, 2013).

3.2.4. MDiagnose

A typical speed test was applied that consisted of easy items which—with unlimited time—almost all mathematics teachers would have been able to solve (Joint Committee, 1999). The task was to diagnose student errors on the level of school mathematics within rather limited time (so-called “experimenter-paced testing”; Davison, Semmes, Huang, & Close, 2012). In this setting, the test-taker is informed beforehand that there will be a very short time-limit within which the response has to be given.

16 items representing typical misconceptions of lower-secondary students (e.g., Malle, 1993) were used (König et al., 2014; Pankow et al., in press). Before three different student solutions were shown on the screen, information about the respective mathematical area was given (e.g., multiplying fractions), and the teachers were asked to anticipate typical student errors. They could decide themselves when they wanted to see the three student solutions but once they had done so, they had to diagnose the error within four seconds. If they did not press any button within this time, the item was coded as wrong. If they defined a correct student solution as incorrect, the item was coded as wrong as well. Thus, information under time pressure was combined with accuracy so that the final score was a function of both (Arthur, Doverspike & Bell, 2004).

3.2.5. P-I-D CM and P-I-D math

To capture situation-specific perceptual, interpretation and decision-making skills, a video-based assessment was developed that required teachers to perceive typical classroom situations presented in three short video clips (Kaiser et al., 2015). Some incidents were presented only very briefly or were not in the center of a clip. The teachers had to rate what they had *perceived* from a mathematics instruction or a general classroom management point of view on a 4-point Likert scale (Clausen, Reusser, & Klieme, 2003). 22 ratings constituted general classroom management skills as part of the P-I-D CM scale, 16 ratings constituted mathematical teaching skills as part of the P-I-D math scale (called P_PID or M_PID, respectively, in Kaiser et al., 2015).

26 experts, including university and school-based practical experts as well as experts from mathematics and general pedagogy, confirmed the content validity of the statements and decided which rating could be accepted as correct. These ratings were used as benchmarks for classifying the teachers' responses. Items were accepted for the main study if at least 60% of the experts agreed on one of the four specifications of the Likert scale (1 = “fully correct”, 2 = “partially correct”, 3 = “partially not correct”, 4 = “not correct at all”). If this agreement was not reached, but a minimum of 80% agreed on the tendency (combining 1 and 2 on the one hand and 3 and 4 on the other hand), the item was revised for the main study, incorporating the explanatory notes of the experts to optimize the test instrument. Items that did not reach a consent were eliminated. The item modification was followed by a second expert rating in which the final agreement was 85%.

In addition to these ratings, teachers were asked to analyze and *interpret* the classroom situations shown in the three videos from a general pedagogical and a mathematical teaching perspective and to make *decisions* about mathematical or pedagogical strategies. 18 of these open-ended items belonged to the P-I-D CM and 18 to the P-I-D math scale. Again, expert panels confirmed twice the content validity of the instrument, and this prior to the pilot study and prior to the main study. The experts discussed the relevance and the frequency of the classroom situations presented in the three videos and the specific items. Since the assessment covered general pedagogical and mathematical classroom demands, experts from both fields and with university as well as practical backgrounds were involved.

Several pilot studies with student teachers, beginning teachers and experienced teachers further confirmed the construct validity of the video assessment. A detailed coding manual and trainings ensured high inter-rater reliability ($\kappa = 0.86$, min–max = 0.76–1.00; Jonsson & Svingby, 2007). The final reliability accomplished was moderate to good (see Table 2), in line with the reliability typically accomplished in performance assessments (see, for example, Kersting, 2008; Gold et al., 2013) and sufficient for analyses on the group level.

3.3. Example of the video-based assessment

The three video clips which served as cues lasted between 2.5 and 4 min and showed lower-secondary mathematics instruction in Germany. They were scripted based on the state of research and displayed characteristics of expertise (Kaiser et al., 2015). The assessments were built with the CBA ItemBuilder, which is a graphical authoring system for complex item development (Rölke, 2012). Before the teachers watched the clips, they were informed in detail about the context of the lessons. Every clip could only be watched once.

One example is a video clip about the volume of an open-top box folded with a standard A 4 piece of paper (corresponding roughly to letter size). The clip lasts for 3.5 min and displays a lesson in grade 9 at the academic track of lower-secondary school (*Gymnasium*). The students work in pairs to find the volume of the box. Depending on the way the box is manually or mentally built by the students, different outcomes of the volume are possible. The clip shows three pairs working together in different ways from a mathematical (P-I-D math) and a general pedagogical point of view (P-I-D CM).

The teachers' *perceptual* skills were assessed with ratings of statements such as “The problem presented by the teacher

Table 3
Fit of the models hypothesized to the data and parameter estimates.

	Fit indices			χ^2 tests		Standardized factor loadings					
Model	CFI	RMSEA	SRMR	χ^2	χ^2/df	MCK	MPCK	GPK	P-I-D CM	P-I-D math	MDiagnose
2-dim «generalizability» (H1)	1.00	0.01	0.05	8.1^{ns}	1.0^{ns}	0.96*	0.74*	0.54*	0.54*	0.99*	0.48*
2-dim «domain-specific» (H2)	0.60	0.23	0.12	77.2*	9.6*	0.86*	0.82*	0.50*	0.56*	0.60*	0.46*
2-dim «method-specific» (H3)	0.72	0.19	0.11	55.7*	7.0*	0.83*	0.85*	ns	0.59*	0.89*	ns
1-dim «homogeneity» (H4)	0.60	0.21	0.12	78.0*	8.7*	0.85*	0.83*	0.37*	0.44*	0.61*	0.46*

Note. CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMS = standardized root mean square residual, df = degrees of freedom, MCK = mathematics content knowledge, MPCK = mathematics pedagogical content knowledge, P-I-D = perception, interpretation and decision-making skills, MDiagnose = Skill to diagnose mathematical student errors, *p < 0.01, ns = not significant.

contained typical characteristics of open word problems" (P-I-D math), or "Most students participated actively in the lesson" (P-I-D CM). They had to *interpret* the classroom situation based on questions such as "The three pairs chose different approaches to solve the mathematical problem. Please briefly describe the core aspects of the approaches. If possible, use the corresponding technical terms" (P-I-D math). The general pedagogical perspective (P-I-D CM) was covered by questions such as: "Please describe the type of seat work the students used to solve the problem." The teachers also had to make *decisions* with respect to further instructional steps. This was prompted by questions such as "The teacher intends to initiate a discussion of the different group results as a whole-class activity by closing the seat work with the remark 'Okay. What are your solutions?' Please describe two alternatives for how to ask the students to exchange their results in a less teacher-centered way." (P-I-D CM).

All instruments were delivered online. Blocks of items were timed so that control existed to prevent cheating. To avoid an increase in item difficulties due to fatigue, a rotated test design was applied with eight blocks of items.

4. Scaling and data analysis

In a first step, scaled scores for each of the six facets of teacher cognition were created. The Rasch model and robust estimators that can deal with non-normal distribution of data were applied with the software Conquest (Wu, Adams, & Wilson, 1997). Items omitted or not reached were treated as wrong. In case of MCK, MPCK and GPK, item parameters available from TEDS-M were imported.

In a second step, the hypotheses were tested by Confirmatory Factor Analysis (CFA) using the six scale scores as indicators. Maximum likelihood estimation and a chi-square test statistic robust to non-normality were applied (MLR). Missing data that had occurred because some teachers had skipped one of the assessments was handled by applying the full information maximum likelihood (FIML) procedure implemented in MPlus 5.1. The model fit was evaluated with the Comparative Fit Index (CFI), two global fit indices (Root Mean Square Error of Approximation, RMSEA, and Standardized Root Mean Square Residual, SRMR), the Chi-square test as well as the ratio of the Chi-square deviance and the degrees of freedom. CFI estimates >0.95 indicate a very good, estimates >0.90 a good model fit (Hu & Bentler, 1999). RMSEA and SRMR estimates <0.05 indicate a very good, estimates <0.08 a good model fit. χ^2/df estimates <2 are regarded as a very good and <3 a good fit.

5. Results

5.1. Distinction of cognitive facets according to level of generalizability (H1)

The model that hypothesized different levels of generalizability as the crucial distinction between facets of teacher cognition revealed a very good fit to the data (see Table 3), and this with respect to relative (CFI = 1.0) and absolute fit indices (RMSEA = 0.01; SRMR = 0.5) but also with respect to the chi-square tests ($\chi^2/df = 1.0$). H1a was therefore strongly supported by the data. As hypothesized (H1c), the first dimension included MCK, MPCK and MDiagnose as rather stable indicators generalizable across mathematics teaching situations whereas the second dimension included the more situation specific indicators P-I-D math, P-I-D CM and GPK (H1d). The latent correlation between the two dimensions was – as hypothesized in H1b – significantly positive ($r = 0.41$).

The variance of all six indicators was significantly—and in two cases almost completely (P-I-D math and MCK)—explained by the underlying respective dimensions. However, although each time significant, the loadings of the different indicators varied. MDiagnose ($\lambda = 0.48$), GPK ($\lambda = 0.54$) and P-I-D CM ($\lambda = 0.54$) correlated lower with the underlying dimension than the other cognitive facets.

5.2. Competing models (H2–H4)

All three competing models, distinguishing the cognitive facets according to domains (mathematics vs. pedagogy, H2) or assessment format (digitalized paper-pencil vs. video-based; H3) or claiming homogeneity (H4), fit worse to the data than the generalizability model (see Table 3). Neither the relative (CFI = 0.60–0.72) nor the absolute fit indices (RMSEA = 0.19–0.23; SRMR = 0.11–0.12) of the competing models were close to acceptable thresholds of CFI ≥ 0.90 or RMSEA/SRMR ≤ 0.08 respectively. The ratios of chi-square values to degrees of freedom were each time very large ($\chi^2/df = 7.0–9.6$) and did not support any of the models, either.

In each of the competing models, the variance of at least one indicator was not significantly explained by the underlying dimension. In the model that distinguished between assessment formats, the factor loadings of two indicators – GPK and MDiagnose – were not significant (see Table 3). In general, several indicators had substantially lower factor loadings in the competing models than in the generalizability model. This applied particularly to MCK, GPK and P-I-D math. Thus, the data did neither support the hypothesis that teacher cognition has to be differentiated into the two domains mathematics instruction and general pedagogy (H2), according to the two assessment formats digitalized paper-and-pencil test and video-based assessment (H3) nor that teacher cognition is homogenous at this stage of a teacher's career (H4).

5.3. Relation of teacher education to teachers' cognitive dimensions (H5)

The differences in means (H5a) and the differences in relations between the two dimensions (H5b) for lower-secondary mathematics teachers and those also teaching in upper-secondary school reflected the hypothesized group differences. The stable and the situation-specific cognitive dimensions were much stronger related to each other in the upper-secondary group ($r = 0.72^*$) than in the lower-secondary group where the relation was not even significant ($r = 0.30$). This means that the cognitive base of the first group was better integrated because the two dimensions were well-connected.

Furthermore, whereas the two teacher groups did not differ significantly on their mean level of situation-specific cognitions, the difference in their stable cognitions was substantial ($M = 0.00$ or $M = 1.44$). The difference corresponded to almost 1.5 standard deviations of the score variance which was not only a highly significant but also a substantial difference. This measure of dispersion shows that the overlap in scoring between the two teacher groups was limited.

The data supported, finally, that the strength of the relation between the grade in the practical teaching exam and teacher cognitions differed by dimension as hypothesized (H5c). The relation to the situation-specific dimension was significant ($r = 0.22^*$) whereas the relation to the stable dimension was lower and not significant ($r = 0.03$).

6. Discussion

A two-dimensional model of teacher cognition distinguishing between levels of generalizability across different situations of mathematics teaching was strongly supported by test data from 171 German mathematics teachers in their third year in the profession. Data from direct assessments of six cognitive knowledge and skill facets showed a better model fit than competing models that distinguished these facets according to domains or assessment formats or claimed homogeneity. MCK, MPCK and MDiagnose on the one hand as well as P-I-D math, P-I-D CM and GPK on the other hand were grouped together as distinct cognitive dimensions. The first one can be described as a rather stable cognitive base teachers can draw on while they plan or reflect on different types of mathematic-specific or general pedagogical teaching activities. In contrast, the second dimension included cognitive facets which are more variable and related to specific characteristics of enacted teaching situations in the classroom.

The two dimensions represent characteristics that Swanson, O'Connor, and Cooney (1990) had identified as mental characteristics specific to the information processing of expert teachers but they challenge the familiar distinction of knowledge on the one hand and skills on the other hand. Furthermore, the results challenge the assumption that content knowledge, pedagogical content knowledge and general pedagogical knowledge are of the same nature (Blömeke, Kaiser & Lehmann, 2010; Baumert et al., 2010; Shulman, 1987; Schoenfeld & Kilpatrick, 2008). It seems instead to be necessary and more fruitful to elaborate always in more detail on the nature of such constructs, no matter whether a theoretical or an empirical study is reported. Depending on the concrete definition and operationalization, it may be that a knowledge as well as a skill facet reflect either a more stable or a more variable cognition. This may particularly apply to MPCK which is sandwiched between mathematics and general pedagogy (as emphasized and conceptualized by Shulman, 1987). Depending on how precisely the construct is defined in a theory or in an assessment, it may be more closely linked to one or the other. In TEDS-M, the relation to MCK was dominant (Döhrmann et al., 2012).

A two-dimensional domain-specific, a two-dimensional method-specific and a one-dimensional homogeneity model could clearly be rejected. This result means, firstly, that domain-specificity should in case of teachers not be defined too narrowly as it is typically done in expertise research. P-I-D math was, for example, more closely related to P-I-D CM and GPK than to MCK. The rejection of method-specificity as major distinction revealed that a controversy about "appropriate" approaches to the assessment of knowledge and skills may be too superficial. To contrast paper-and-pencil tests and video-based assessments may overshadow the usefulness of efforts to strive for more complexity of paper-and-pencil tests. Including substantial proportions of open-ended questions requires to think ahead and to generate strategies as it was done in the TEDS-M GPK test (König & Blömeke, 2009).

The results also imply that teacher cognition even after up to 10 years of training and practical experience is not a homogenous but a heterogeneous construct. This result is noteworthy because teacher education in Germany is long and extensive. Particularly the induction period is characterized by elements described as deliberate practice by Ericsson (2005), namely, striving for improvement, receiving high-quality feedback by master teachers, and opportunities to repeat their performance.

The results about the relation between teacher education and the cognitive dimensions show clear patterns which helps us to better understand the trajectories of teacher learning. Stronger prerequisites and much more mathematics-related opportunities to learn during teacher education played favorably out in case of teachers who were prepared to teach upper-secondary mathematics in addition to lower-secondary mathematics. Their stable cognitive base was stronger and it was better connected to situation-specific skills. The investment made by the prospective teachers in terms of time invested in education and by the federal states in terms of opportunities to learn provided paid off. Secondly, no matter which programs teachers had taken, those with stronger situation-specific skills showed classroom performance of higher quality as indicated by grades in the practical teaching exam than teachers with weaker cognitive skills.

6.1. Challenges and limitations of our study

Before conclusions can be drawn from these results, conceptual and methodological limitations have to be pointed out. Although the hypothesized two-dimensional generalizability model fit well to the data, this is still a simplified model of classroom complexity. It includes only two of the many challenges mathematics teachers have to deal with, namely mathematics instruction and classroom management, although these challenges are without doubt particularly relevant for instructional quality and student achievement (Fauth et al., 2014). Furthermore, while each dimension explained a large proportion of variance in one or two indicators, the proportion of variance explained was lower for the other ones. This result may reflect remaining inherent multi-dimensionality of the current dimensions.

With respect to methodological limitations, it has to be pointed out that the study was based on a sufficient sample size given that all hypothesized models involved scaled scores only but not latent variable modeling which was not feasible given the sample size. However, it is technically possible that a model with six latent variables including all 188 items and estimating item and person parameters at the same time revealed a different internal structure of teacher expertise than presented in this paper. Replication with larger sample sizes is therefore much needed. Note also that both constructs, knowledge as well as skills, were characterized as including a certain proportion of stability or variability respectively which means that the specific nature of a facet depends on its concrete definition and operationalization. Furthermore, the sample was restricted to teachers from Germany. The results would be strengthened if replication studies in other countries take place.

7. Conclusions

The distinction of two dimensions and their differential relation to classroom performance point to the necessity to train both dimensions carefully and tailored to teachers' strength and weaknesses. This conclusion applies to initial teacher education but also – and may be even more strongly given the closer link to practicing teachers – to professional development activities. In this context, the low relation between stable and situation-specific cognitions together with the low mean on the first dimension of lower-secondary teachers compared to upper-secondary teachers is the most worrying result of this study. It points to serious disadvantages of the first group which can be related back to fewer opportunities to learn during teachers' university education. Educational policy is prompted to take action here – and this not only in Germany. In many countries such differences in teacher education programs may exist and it is very probable that they result in different outcomes, too. It seems to pay off to offer rich mathematics-specific opportunities to learn so that knowledge and skills can be developed.

Promising research directions can be derived from this study as well. The role of other teacher resources, for example beliefs, motivation or volition has not yet been examined although regarded important for performance in expertise research (Gobet, 2005). Furthermore, an extension beyond the current range of practical teaching experience is desirable. Would a study with mathematics teachers who have, for example, 20 or 30 years of practical experience show the same cognitive structure or would it be, for example, more homogenous?

Finally, whereas we were able to provide evidence that situation-specific cognitions are related to teacher behavior in terms of instructional process in the classroom (indicated by the grade point average in the second state exam based on direct observations of the teachers), the link to student achievement – although plausible – is missing. A study that links teacher cognition, instructional quality and student achievement would be timely.

Acknowledgment

This study was funded by the German Research Foundation (DFG BL 548/8-1). The views expressed are those of the authors and do not necessarily reflect the views of the DFG.

References

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
Arthur, W., Doverspike, D., & Bell, S. T. (2004). Information processing tests. In M. Hersen, & J. C. Thomas (Eds.), *Industrial/organizational assessment: vol. 4. The comprehensive handbook of psychological assessment* (pp. 56–74). New York: Wiley.
Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.). (2010). *TEDS-M 2008: Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*, 3–13.
Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., …, & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom and student progress. *American Educational Research Journal, 47*, 133–180.
Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens*. Bern: Huber.
Clark, C., & Lampert, M. (1986). The study of teacher thinking: implications for teacher education. *Journal of Teacher Education, 37*(5), 27–31.
Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen: Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft, 31*, 122–141.
Davison, M. L., Semmes, R., Huang, L., & Close, C. N. (2012). On the reliability and validity of a numerical reasoning speed dimension derived from response times collected in computerized testing. *Educational and Psychological Measurement, 72*, 245–263.
De Groot, A. D. (1978). *Thought and choice and chess (1946)*. The Hague, Netherlands: Mouton.
Döhrmann, M., Kaiser, G., & Blömeke, S. (2012). The conceptualisation of mathematics competencies in the international teacher education study TEDS-M. *ZDM – The International Journal on Mathematics Education, 44*, 325–340.
Ericsson, K. A. (2005). Recent advances in expertise research: a commentary on the contributions to the special issue. *Applied Cognitive Psychology, 19*, 233–241.
Evertson, C. M., & Weinstein, C. S. (Eds.). (2009). *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 407–437). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9.
Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum (pp. xv–xxvii).
Gobet, F. (2005). Chunking models of expertise: implications for education. *Applied Cognitive Psychology, 19*, 183–204.
Gold, B., Förster, St., & Holodynski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionallen Wahrnehmung von Klassenführung im Grundschulunterricht. *Zeitschrift für Pädagogische Psychologie, 27*, 141–155.
Hill, H. C., Rowan, B., & Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371–406.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
Joint Committee on Standards for Educational and Psychological Testing of the AERA, APA, and NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review, 2*, 130–144.
Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the complexities of video-based assessments: Theoretical and methodological approaches to overcoming shortcomings of research on teachers' competence. *International Journal of Science and Mathematics Education, 13*, 369–387.
Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement, 68*, 845–861.
KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2006). *Abiturnoten an allgemein bildenden Gymnasien und integrierten Gesamtschulen: Schuljahr 2004/05*. Retreived from: http://www.gew.de/Binaries/Binary29527/5KMK-Abiturnoten_2005%20%282%29.pdf [20.08.14].
König, J., & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften: Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerausbildung. *Zeitschrift für Erziehungswissenschaft, 12*(3), 499–527.
König, J., Blömeke, S., 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.
Krauss, S., & Brunner, M. (2011). Schnelles Beurteilen von Schülerantworten: Ein Reaktionszeittest für Mathematiklehrer/innen. *Journal für Mathematikdidaktik, 32*, 233–251.
Malle, G. (1993). *Didaktische Probleme der elementaren Algebra*. Braunschweig, Wiesbaden: Vieweg.
Pankow, L., Kaiser, G., Busse, A., Hoth, J., Döhrmann, M., König, J. & Blömeke, S. (in press). Early Career Teachers' Competence to Identify Student Errors: Results from an Innovative Test Approach. In press in *ZDM - The International Journal of Mathematics Education*.
Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.
Rölke, H. (2012). The ItemBuilder: a graphical authoring system for complex item development. In T. Bastiaens, & G. Marks (Eds.), *Proceedings of world conference on e-learning in corporate, government, healthcare, and higher education 2012* (pp. 344–353). Chesapeake, VA: AACE.
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*, 63–88.
Schoenfeld, A. H. (2010). *How we think: A theory of goal-oriented decision making and its educational applications*. New York: Routledge.
Schoenfeld, A. H., & Kilpatrick, J. (2008). Toward a theory of proficiency in teaching mathematics. In D. Tirosh, & T. Wood (Eds.), *Tools and processes in mathematics teacher education: vol. 2. International handbook of mathematics teacher education* (pp. 321–354). Rotterdam, Netherlands: Sense Publishers.
Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York: Routledge.
Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review, 57*, 1–22.
Smith, T. W., & Strahan, D. (2004). Toward a prototype of expertise in teaching: a descriptive case study. *Journal of Teacher Education, 55*, 357–371.
Star, J. R., & Strickland, S. K. (2008). Learning to observe: using video to improve

preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education, 11*, 107–125.

Stürmer, K., Könings, K. D., & Seidel, T. (2012). Declarative knowledge and professional vision in teacher education: effect of courses in teaching and learning. *British Journal of Educational Psychology, 83*, 467–483.

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.

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. L., Ingvarson, L., Rowley, G., …, & Reckase, M. (2012). *Policy, practice, and readyness to teach primary and secondary mathematics in 17 Countries: Findings from the IEA teacher and development study in mathematics (TEDS-M)*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Van Overschelde, J. P., Rawson, K. A., Dunlosky, J., & Hunt, R. R. (2005). Distinctive processing underlies skilled memory. *Psychological Science, 16*, 358–361.

Wu, M. L., Adams, R. J., & Wilson, M. (1997). *ConQuest: Generalised item response modelling software*. Hawthorn, Australia: ACER.