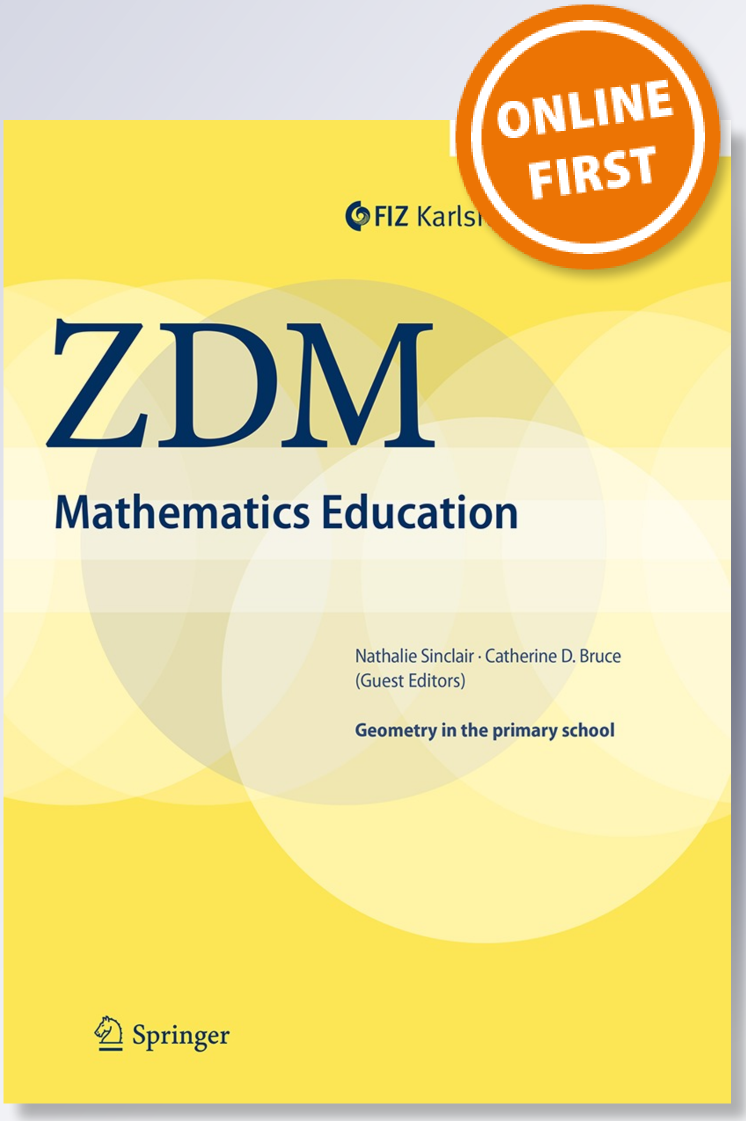
Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME)

Johannes König & Charlotte Kramer

ZDM
Mathematics Education

ISSN 1863-9690

ZDM Mathematics Education
DOI 10.1007/s11858-015-0705-4



Springer

ZDM Mathematics Education
DOI 10.1007/s11858-015-0705-4

ORIGINAL ARTICLE

Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME)

Johannes König[1] · **Charlotte Kramer**[1]

Accepted: 14 June 2015

Abstract Due to the need for measurement instruments that allow an investigation of teachers' situational cognition and thus go beyond the limited scope of classical paper-and-pencil-tests, we ask how a specific video-based measurement of teachers' classroom management expertise can provide additional information when compared with an established paper-and-pencil-test that broadly covers mathematics teachers' general pedagogical knowledge. For this, we apply the general pedagogical knowledge test previously developed in the *Teacher Education and Development Study—Mathematics* (TEDS-M) comprising knowledge of structuring lessons ('structure'); motivating students and managing the classroom ('motivation/classroom management'); dealing with heterogeneous learning groups ('adaptivity'); and assessing students ('assessment'). Using test data of 188 novice teachers, advanced beginners, and expert teachers, we raise questions regarding the two tests' (1) structural relations, (2) expert-novice differences, and (3) predictive validity. Findings: (1a) classroom management expertise can be empirically separated from general pedagogical knowledge, although the two constructs are positively inter-correlated (medium effect size), (1b) classroom management expertise is more highly correlated with pedagogical knowledge of classroom management than with pedagogical knowledge of 'adaptivity', 'structure', and 'assessment', (1c) classroom management expertise is more highly correlated with procedural pedagogical knowledge (cognitive demand 'generate') than with declarative pedagogical knowledge (cognitive demands 'recall' and 'understand/analyze'), (2) novice teachers as well as advanced beginners are outperformed by expert teachers, and (3) classroom management expertise, compared with general pedagogical knowledge, is a stronger predictor for instructional quality aspects of classroom management as rated by students.

Keywords Classroom management · General pedagogical knowledge · Teacher expertise · Video clips · Test · Assessment · Instructional quality

1 Introduction

For the past decades, the interest in doing research on the measurement of cognitive elements of teacher competence has been growing (Blömeke and Delaney 2012; König 2014) due to the understanding knowledge is required for effective teaching. For the majority of relevant studies, however, the classical paper-and-pencil assessment represents the dominating paradigm (e.g., Hill et al. 2008; Tatto et al. 2012) not least because it enables an efficient and reliable way to measure declarative-conceptual knowledge in large samples.

For example, in 2008 the *Teacher Education and Development Study—Mathematics* (TEDS-M) was carried out, a comparative study of teacher education (Tatto et al. 2012). The TEDS-M target population were mathematics teachers for elementary and middle schools in their final year of teacher education. More than 20,000 future teachers from 17 countries worldwide were tested using paper–pencil instruments measuring their mathematical content

Electronic supplementary material The online version of this article (doi:10.1007/s11858-015-0705-4) contains supplementary material, which is available to authorized users.

✉ Johannes König
johannes.koenig@uni-koeln.de

[1] Empirical School Research, University of Cologne, Gronewaldstr. 2, 50931 Cologne, Germany

Published online: 27 June 2015

knowledge, mathematical pedagogical content knowledge, and—due to feasibility in three countries only—their general pedagogical knowledge. Thus, three domains of teacher knowledge were distinguished following the current state of research on teacher knowledge and teacher competence (Shulman 1987; Tatto et al. 2012; Blömeke et al. 2015). In TEDS-M, mathematical content knowledge covers the main mathematical areas relevant for future teachers, whereas mathematical pedagogical content knowledge refers to curricular knowledge, knowledge of lesson planning, and interactive knowledge applied to teaching situations (Döhrmann et al. 2012). General pedagogical knowledge is structured in a task-based way, i.e., referring to knowledge teachers need to prepare, structure and evaluate lessons ('structure'), to motivate and support students as well as manage the classroom ('motivation/classroom management'), to deal with heterogeneous learning groups in the classroom ('adaptivity'), and to assess students ('assessment') (König et al. 2011). The basic assumption underlying TEDS-M was that mathematics teachers need to draw on this range of knowledge in order to be able to provide high quality opportunities to learn to their students. Although general pedagogical knowledge is not subject-specific, nevertheless it constitutes an equally important cognitive element of mathematics teachers' professional competence (Shulman 1987; Bromme 1992).

However, the measurement of context-dependent, procedural teacher knowledge goes beyond the limited scope of classical paper-and-pencil assessments (Shavelson 2010; Blömeke et al. 2015). There is the need for instruments that allow an investigation of teachers' situational cognition and the impact of individual differences in teaching experience and in-school opportunities to learn during teacher education (König et al. 2014). Although knowledge acquired during teacher education and represented as declarative knowledge is probably of great significance, especially the research on teacher expertise has worked out that both declarative and procedural knowledge contributes to the expert's performance in the classroom (Bromme 2001).

To account for such methodological concerns a major current focus in the measurement of teacher knowledge and skills is the shift from paper-and-pencil tests to the implementation of instruments using video clips of classroom instruction as item prompts: such studies use videos as a stimulus in the item stem, an assessment format which is frequently referred to as "video-vignette" or "video-cued testing". Video-based assessment instruments are used to address the contextual nature and the complexity of the classroom situation. They are considered to improve the measurement of teacher knowledge when compared with the classical paper-and-pencil test (Kaiser et al. 2015; König 2015a, 2015b).

Several studies adopted this approach to provide a more ecologically valid measurement of the knowledge of mathematics teachers (e.g., Kersting 2008; König et al. 2014) or the knowledge of teachers in general (e.g., Seidel et al. 2010; Voss et al. 2011; Gold et al. 2013). These studies thus intend to measure knowledge that is more of a situated nature (Putnam and Borko 2000). With the growing popularity of video-based measurements in the field of teacher knowledge research, it is essential to establish a convincing empirical rationale for their implementation. However, the theoretical and methodological advantage delivered by using video clips remains to be specified. To expand previous research, our study aims to address the measurement of situational knowledge in teachers by proposing a video-based approach for testing pedagogical knowledge and skills required for successfully meeting the specific requirements involved in effective classroom management. To do so, we build our study on previous research on the measurement of general pedagogical knowledge that was conducted in TEDS-M (König et al. 2011). The instrument measuring teachers' general pedagogical knowledge that was developed in TEDS-M is applied in our study as well. Against the background of this instrument that captures general pedagogical knowledge by using a broad range of content, we ask how a more specific measurement of teachers' classroom management expertise using a novel video-based assessment can provide additional information to describe and analyze teachers' professional knowledge in the field of general pedagogy. We test the general pedagogical knowledge and classroom management expertise of teachers who are at various stages during their expertise development (cf. Berliner 2001)—'novice teachers' (i.e., student teachers at the start of initial teacher education), 'advanced beginners' (i.e., teacher candidates during their practical training at the end of initial teacher education), and 'expert teachers' (i.e., in-service teachers with, by average, 18 years teaching experience)—and conduct three kinds of analyses: (1) construct validation analysis, (2) analysis of expert-novice-differences, and (3) analysis of predictive validity. Against the background of our findings, we will discuss to what extent a specific video-based assessment of classroom management expertise is of additional value when compared with a classical paper-and-pencil general pedagogical knowledge test.

2 State of research and theoretical framework

2.1 Defining general pedagogical knowledge in TEDS-M

Although consensus exists towards the definition which Shulman (1987, p. 8) provided, namely that general pedagogical knowledge involves "broad principles and strategies of classroom management and organization that appear to transcend subject matter" as well as knowledge about learners and learning, assessment, and educational contexts and purposes, there was a lack of empirical studies on teachers' general pedagogical knowledge (Wilson and Berne 1999) when TEDS-M started. Virtually no studies existed that could have shown how to fill these relatively broad domains of general pedagogical knowledge outlined by Shulman (1987) so that one could develop items and actually test teachers. Previous to TEDS-M, few studies had approached specific aspects only (see the review by König, 2014). In a joint effort, the German, US and Taiwan TEDS-M teams developed therefore a theoretical framework of teachers' general pedagogical knowledge that could be transformed into a paper-and-pencil instrument and be tested empirically across countries. Following the notion of "competence" (Shavelson 2010; Weinert 2001; specified for the teaching profession by Bromme 1992, 2001), the study's framework focused on the mastering of professional tasks and its underlying latent cognitive dispositions. This meant that the theoretical framework of general pedagogical knowledge was structured in a task-based way and explicitly not according to the formal structure of general pedagogy as an academic discipline.

Instruction was identified as the core activity of teachers in all subjects and countries (Berliner 2001; Bromme 1992). Against the background of instructional models used across countries to describe effective teaching (Slavin 1994; Good and Brophy 2007) the theoretical framework was defined to consist of four generic dimensions of teaching quality (see Supplementary Material, Figure 1; for more details see König et al. 2011): Thus teachers are expected to have general pedagogical knowledge allowing them to prepare, structure, and evaluate lessons ('structure'); to motivate and support student learning as well as to manage the classroom ('motivation/classroom management'); to deal with heterogeneous learning groups in the classroom ('adaptivity'), and to assess students ('assessment'). In the model of professional competence underlying a study like TEDS-M, such knowledge is distinct from mathematical pedagogical content knowledge, but nevertheless is also considered to be essential for teachers of mathematics.

In addition, three cognitive demands on teachers when dealing with such generic classroom situations were defined following Anderson and Krathwohl (2001): to recall information from long-term memory in order to describe the classroom situation ('recall'); to understand or analyze a concept, a specific term or a phenomenon outlined ('understand/analyze'); and to generate strategies for how they would solve the problem posed ('generate'; for more details, see König et al. 2011). Generic dimensions of teaching quality and cognitive demands made up a matrix which served as a heuristic for the development of general pedagogical knowledge (see Supplementary Material, Figure 2).

The knowledge tested with these cognitive demands is of different quality. Distinguishing between declarative and procedural knowledge is very common in teacher research (besides Anderson and Krathwohl 2001 see e.g., Bromme 2001). In the TEDS-M instrument measuring general pedagogical knowledge, test items requiring teachers to recall, understand or analyze information predominantly measure declarative knowledge ("knowing that…") including factual and conceptual knowledge while test items requiring teachers to generate strategies not only measure declarative but also procedural knowledge ("knowing how…"). Procedural knowledge is of a situated nature (Putnam and Borko 2000).

2.2 Findings on teachers' general pedagogical knowledge

In TEDS-M 2008, the general pedagogical knowledge test was successfully validated through expert reviews in the participating countries and through confirmatory approaches based on large-scale data from these countries (König et al. 2011). Further studies have been carried out to apply the test again using various samples of pre-service and in-service teachers in Germany, but also in other countries (e.g., Austria). All these studies report good psychometric properties of the general pedagogical knowledge test. Reliability of the overall test score is good, e.g., 0.86 for pre-service elementary school teachers (König 2013), 0.78 for pre-service middle school teachers (König et al. 2011), and 0.81 for in-service teachers (König et al. 2014). Besides using an overall test score, the test can be differentiated into sub-dimensions (see Supplementary Material, Figure 2): Reliable test subscales measuring content dimensions (structure, adaptivity, classroom management/motivation, assessment) and cognitive demands (recall, understand/analyze, generate) were created and used to report strengths and weaknesses of a country's performance in TEDS-M (König et al. 2011). In accordance with assumptions of the acquisition of teacher expertise (Berliner 2001), in-service teachers outperform pre-service teachers who are at the end of their initial teacher education, whereas they in turn outperform future teachers just entering initial teacher education (König 2013).

2.3 Relating teachers' general pedagogical knowledge to classroom management knowledge and skills

As TEDS-M shows, teacher knowledge about classroom management can be assigned to a broader understanding of general pedagogical knowledge, whereas in turn general pedagogical knowledge has been defined as one of the central cognitive components of professional teacher competence (Tatto et al. 2012; Blömeke et al. 2015). Since such a conceptualization of teacher competence is based on the research on teacher expertise (König 2014), in the following we consider that classroom management expertise, i.e., the teacher's specific knowledge and skills related to the challenge of managing a classroom, belongs to the area of general pedagogical knowledge thus contributing to an essential component of professional teacher competence.

Research on classroom management in general has triggered broad interest, as for example the corresponding handbook by Evertson and Weinstein (2006) demonstrates. Meta-analyses of empirical studies have repeatedly shown adequate mastery of classroom management is clearly related to student achievement (e.g., Hattie 2012). Successful classroom management depends on the teachers' ability to identify and interpret the critical aspects of the teaching learning process (Kounin 1970). Knowledge in relation to classroom management refers to an "intellectual framework" (Doyle 1985, p. 33), consisting of knowledge of the learning environment and procedures for adequate classroom management, which teachers have to acquire rather than an accumulation of isolated scripts and facts such as "don't smile before Christmas".

As can be seen with TEDS-M, the importance of classroom management expertise as being part of a teacher's professional knowledge has already been addressed by some studies on teacher competence measurement (e.g., Seidel et al. 2010; Voss et al. 2011, see, for details, the systematic literature review by König 2014). Although these studies seize classroom management as an important aspect of teachers' general pedagogical knowledge, none of them have started to conceptualize teachers' situational knowledge of classroom management extensively. As a consequence, the measurement of classroom management has not only been limited to the paper-and-pencil approach predominantly, but it has also been kept a subordinated construct of general pedagogical knowledge. This becomes critical when a measure of classroom management expertise conceptualized as a self-contained construct is needed, for instance, for doing specific research on the effectiveness of professional development of teachers in the field of classroom management. Also, the question arises whether classroom management expertise might function as a predictor for aspects of instructional quality related to effective classroom management that is more proximal than general pedagogical knowledge. Since in general there is hardly any study that investigates how the knowledge of teachers is related to the instructional quality of their teaching, in the following this will be examined.

2.4 Conceptual framework of teachers' Classroom Management Expertise (CME)

To establish the theoretical rationale for our study, a specification of pedagogical knowledge and skills required for successful classroom management was developed (cf. König and Lebens 2012). Building on previous research showing expert teachers systematically perceive and interpret classroom events and sequences differently from novices, three cognitive demands that will be outlined in the following were distinguished: 'accuracy of perception', 'holistic perception', and 'interpretation/justification of action'.

First, from the research on teacher expertise which has proven to be valid across different subjects and countries, it is well known that expert teachers outperform novice teachers in recalling meaningful instructional details (König and Lebens 2012). Expert teachers' categorical perception with which phenomena, events, or sequences are cognitively divided into relevant units for perception (e.g., Bromme 1992) supports them to focus on the *relation* between knowledge elements rather than on discrete elements. Repeated activation of schemata strengthens connections between elements within a schema and support enhanced activation of knowledge for categorizing new information when salient cues are present. Since connectivity and complexity of schemata required for identifying and categorizing information evolve with practice (Berliner 2001), 'accuracy of perception' is an indicator of expertise. Consequently, it can be reasonably assumed that expert teachers identify relevant instructional situations seen in a video-vignette assessment more precisely and correctly than do novices (Sabers et al. 1991).

Second, expert teachers can be characterized by their 'holistic perception', especially when compared with novice teachers (Bromme 2001; König and Lebens 2012): expert teachers reconstruct and anticipate the context of instruction and engage in reflecting alternative problem-solving strategies. Whereas novices slowly observe classroom situations step by step due to the fragmented structure of their knowledge, experts have an intuitive grasp of the situation since their knowledge is highly interlinked (Bromme 1992). More specifically, prior knowledge of experts organized in schemata is employed during perception to form a cognitive representation of the situation. By contrast, novices, whose knowledge structures for constructing a mental framework have not yet been developed, are likely to experience difficulties in reconstructing the context of instruction.

The third dimension of cognitive demands ('interpretation/justification of action') refers to the functional interpretation of instructional events and sequences that depends on reasoning about the instructional intention and rationale amidst the context of classroom teacher-student interaction (Berliner 1992). Although the functional interpretation of actions is rarely explicated in everyday teaching situations, it can be accessed from long-term memory (Bromme 1992). In contrast to teachers' holistic perception, the interpretation of events goes beyond generating mental representations, since it strongly depends on reframing and transforming knowledge (König et al. 2014), especially when teachers are required to justify a teacher's action in the classroom (König and Lebens 2012). Whereas the holistic perception can be described as a perceptive-representational process, the interpretation of events refers to transformative processes.

Besides these three cognitive demands relevant for measuring classroom management expertise, a variety of typical classroom management situations are needed to assure content-related breadth of the assessment. So the video clips used for the assessment in our study refer to typical classroom management situations in which teachers are heavily challenged (following classifications provided by Hawk and Schmidt 1989; Swartz et al. 1990; Doyle 2006), involving to manage transitions, instructional time, student behavior, and instructional feedback. Although each video can be assigned to one of these situations, they also include aspects of the other situations.

2.5 Context of the study: pre-service and in-service teachers in Germany

Internationally compared, Germany is a country with a specific teacher education structure, since it offers teacher education programs that are spread over two phases, a theoretical and a following practical (König and Blömeke 2013). Future teachers begin their preparation in one of the German universities with 4- or 5-year programs. This first phase contains a great deal of required educational coursework with a heavy emphasis on theory. Future teachers finish university with a degree (Bachelor/Master of education) which is the general requirement for entry into the second phase. Most of the practical teacher preparation is provided in the 1.5-year second phase in special, generally small, institutions operated by state governments. Future teachers work part-time at schools and attend courses in general pedagogy and subject-related pedagogy. The second phase ends with a state examination which comprises the assessment of pre-service teachers' teaching performance. After finishing initial teacher education, teachers are not systematically required to engage in professional development courses (cf. König and Blömeke 2013).

3 Research questions

Starting from the research on the measurement of general pedagogical knowledge in the context of TEDS-M, we ask how a more specific measurement of teachers' classroom management expertise using a video-based assessment format can provide additional information to describe and analyze teachers' professional knowledge in the field of general pedagogy. We examine the following three questions:

(1) How is teachers' general pedagogical knowledge as assessed by the TEDS-M instrument structurally related to the novel classroom management expertise-measure using video-cued testing? To conduct this construct validation analysis, we use the following three hypotheses (abbreviated as H1a, H1b, and H1c in the following):

H1a: Considering convergent validity (Campell and Fiske 1959), an examination of the correlation between classroom management expertise and general pedagogical knowledge is of great interest. Since general pedagogical knowledge involves "broad principles and strategies of classroom management and organization that appear to transcend subject matter" as well as knowledge about learners and learning, assessment, and educational contexts and purposes (Shulman 1987, p. 8), classroom management expertise can be regarded as a construct that is located in the field of general pedagogical knowledge, but due to its specific definition covers a segment of that knowledge only. Thus we assume the classroom management expertise measure is positively correlated (medium effect size) but not identical with general pedagogical knowledge as assessed by the TEDS-M test.

H1b: When differentiating the general pedagogical knowledge test into content dimensions, the classroom management expertise measure should be more highly inter-correlated with the subscale 'classroom management/motivation' than with the other subscales 'adaptivity', 'structure', and 'assessement'.

H1c: When differentiating the general pedagogical knowledge test into dimensions of cognitive demands, the classroom management expertise measure should be more highly inter-correlated with the subscale 'generate', which consists of test items requiring teachers to generate strategies that measure not only declarative but also procedural knowledge (König et al. 2011). Since teachers' procedural knowledge is of a situated nature (Putnam and Borko 2000), particularly the subscale 'generate' should be more closely linked to the classroom management expertise measure than declarative-conceptual knowledge which is measured by test items of the other two dimensions of cognitive demands, i.e., the subscales 'recall' and 'understand/analyze'.

(2) Is classroom management expertise developed during initial teacher education and in the course of professional teaching?

H2a: Against the background of the research on teacher expertise, that has worked out differences between expert teachers and novice teachers we assume that in-service teachers (regarded as 'experts' in our study) outperform pre-service teachers in the classroom management expertise-test.

H2b: Moreover, pre-service teachers of the second phase of initial teacher education, i.e., those with some teaching experience (as outlined in Sect. 2.5), should outperform pre-service teachers without any systematic teaching experience, i.e., those who have not finished their Bachelor of education and thus are in the first phase of teacher education.

(3) How is the classroom management expertise of teachers related to the instructional quality of their teaching?

Research on effective teaching aiming at analyzing components of classroom instruction that influence student learning has brought about a plentitude of findings. Meta-analyses have tried to integrate findings from many studies in order to work out the most important factors (e.g., Hattie 2012). Generally our conceptualization of classroom management expertise is based on the assumption that the instructional quality provided by teachers is not independent from their classroom management expertise as assessed via the novel video-based instrument. Thus we include an examination of the predictive validity of the classroom management expertise-measure.

H3a: We assume a positive correlation between classroom management expertise and specific measures (in our study captured via student ratings) of instructional quality related to classroom management.

H3b: Due to its situational nature we consider classroom management expertise as being a proximal predictor of instructional quality related to classroom management. Thus we assume that the correlation assumed with H3a is higher than the corresponding correlation between general pedagogical knowledge and instructional quality measures.

4 Method

4.1 Samples and data collection

In our study we use three samples of pre-service and in-service teachers from the area of Cologne, Germany: (1) a sample of pre-service teachers who were in their first phase of initial teacher education, (2) a sample of pre-service teachers who were in the second phase of initial teacher education, (3) an expert teacher sample, i.e., in-service teachers with several years of teaching experience. Following stage models of teacher expertise development (Berliner 2001), we describe them as 'novices' (1), 'advanced beginners' (2), and 'experts' (3).

1. The first sample consists of 114 student teachers of the University of Cologne who attended one of three regular seminars on teaching methods in the summer term 2014. Right at the start of the term their classroom management expertise and general pedagogical knowledge was tested online. Participation was obligatory thus ensuring a 100 % participation rate. 90 of them (79 %) are female. By average they are approximately 23 years old (M = 23.21, SD = 3.85).
2. The second sample derives from the study *Longitudinal Survey of Pedagogical Competencies of Student Teachers and Teacher Candidates* (Längsschnittliche Erhebung pädagogischer Kompetenzen von Lehramtsstudierenden und ReferendarInnen; LEK-R).[1] It consists of 40 pre-service teachers who entered the second phase of initial teacher education during the academic year 2013/2014. Their classroom management expertise and general pedagogical knowledge was tested online as well. Additionally, the instructional quality of their teaching delivered to students was assessed via student ratings thus enabling us to analyze the third research question. 30 of them (75 %) are female. By average they are approximately 28 years old (M = 27.45, SD = 3.46).
3. The third sample consists of the whole teaching staff of two schools in the greater area of Cologne. The schools varied with respect to size from 15 teachers to 19 teachers. Thus the sample consists of 34 teachers. 31 of them (91 %) are female. By average they were approximately 43 years old (M = 43.18, SD = 9.53) and had taught for 18 years at school (M = 17.9, SD = 10.4).

When surveyed, teachers first had to complete a background questionnaire containing variables such as age, sex, and teaching experience. Second, the classroom management expertise instrument was provided with a total duration of 20 min allowing 5 min for watching one video clip and responding to the corresponding test items. The four video clips are very short (they vary between 1 and 2 min in length). Each video clip was presented only once, and respondents were only allowed to read test items related to a video clip when they had already watched that clip. This procedure assured that video clips were used as item prompts in a standardized way and teachers had to respond to test items immediately after having watched the

[1] The LEK-R study has been funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, KO3947/3-2).

correspondent clip. Third, teachers had to complete the paper-and-pencil instrument measuring their general pedagogical knowledge, which took another 20 min. Instructional quality delivered by pre-service teachers [sample (2) only] was captured via ratings of their students. Students were given a questionnaire with several items measuring various aspects of classroom management. In the following, survey data of 449 students rating the instructional quality of 21 pre-service teachers will be included in the data analysis to examine our third research question.[2]

4.2 TEDS-M paper-and-pencil-test measuring general pedagogical knowledge

In this study, we applied a short form of the general pedagogical knowledge test that was developed in the context of TEDS-M (König et al. 2011). A selection of test items was used to reduce the test length to 20 min due to data collection constraints. When applied to the sample of our study, as a first step classical item analysis was conducted over the 34 items. Internal consistency was estimated at $\alpha = 0.87$, which is a good result taking into account that only about half the test items of the original instrument were included into this short form.

As laid out above, generic dimensions of teaching responsibilities and cognitive demands made up a matrix which served as a heuristic for the development of general pedagogical knowledge items in TEDS-M (see Supplementary Material, Fig. 2). For each cell, a subset of items was developed (for details of test development, see König et al. 2011). Two item examples (see Supplementary Material, Fig. 3 and Fig. 4) may illustrate the general pedagogical knowledge test and the heuristic used to conceptualize general pedagogical knowledge (for a more detailed description of the test see, e.g., König & Blömeke, 2010; König et al., 2011; König 2014).

4.3 Video-based assessment of classroom management expertise

4.3.1 Instrument design

The novel classroom management expertise measurement instrument consists of four video clips of classroom instruction that refer to typical classroom management situations in which teachers are heavily challenged. These video clips were carefully selected from a pool of video clips available to the research team. For conceptual reasons, the selection procedure applied mainly intended to follow classifications of typical classroom management situations found in the literature (Hawk and Schmidt 1989; Swartz et al. 1990; Doyle 2006): The video clips had to represent authentic and comprehensive situational information of classroom instruction in which a teacher is challenged (1) to manage transitions, (2) to manage instructional time, (3) to manage student behavior, and (4) to manage instructional feedback. Whole-class interaction teaching situations were preferred, as in terms of effective classroom management they are more complex and thus more challenging for teachers than private work time situations during which a teacher assists a single student or a group of students (Kounin 1970). The video clips had to represent a variety of classroom contexts (regarding school grade, school subject, composition of the learning group, age of teacher), not least in order to detain respondents from getting used to one specific situational context during assessment. Besides conceptual issues, technical criteria had to be met, too. The video clips had to be of good quality both visually and acoustically, they had to represent usual events somehow familiar to every experienced school teacher, they had to be short and self-contained for research-related economic reasons. In a pilot study (König and Lebens 2012) these criteria were issued by conducting an expert review before the procedure of selecting appropriate video clips was started. The video clips do not come along with complementary information about the teacher, the learning group or the lesson, since our idea to measure classroom management expertise was to stick as closely as possible to the situation presented via video and to not distract respondents from perceiving the concrete classroom instruction. However, in contrast to the paper-and-pencil test measuring general pedagogical knowledge that has been successfully applied in various cultural contexts (König et al. 2011), we consider the classroom management expertise test to have a far more limited scope since it is both culturally and linguistically tied to the German-speaking context of teaching.

4.3.2 Item development

Test items were developed for each video clip covering the three cognitive demands outlined above (accuracy of perception, holistic perception, interpretation/justification of action). In total, 27 test items were developed. 7 are multiple-choice-response, 20 are open-response items. Accuracy of perception is measured by 15, holistic perception by 8, and justification of action by 4 test items (see item examples in the Supplementary Material, Figs. 5, 6, and 7).[3]

[2] This group of 21 pre-service teachers was selected from sample (2), since they pursued a teaching career for primary and lower secondary level and thus had identical opportunities to learn in the area of general pedagogy during initial teacher education.

[3] Due to the nature of the cognitive demand, justification of action is only measured by open-response-items, whereas the other two demands are measured both by open-response-items and by multiple-choice-items. However, to account for this, we conducted detailed item-analyses which are documented in König (2015b). Findings show there are no substantial effects the different formats have on correlational patterns.

Coding rubrics were developed for the open-response items in a complex and extensive interplay of deductive (from our theoretical framework) and inductive approaches (from empirical teacher responses). In a pilot phase, codes from several independent raters were discussed in detail and coding rubrics were carefully revised and expanded. Thus, the coding manual is theoretically based as well as data-based. The codes were intended to be low-inferent.

Coding rubrics for open-response items consist of one criterion, two criteria or more than two criteria. If the single criterion is met by the response provided by the respondent, then the rater has to code this criterion with 1. If it is not met, a 0 will be given. However, when doing frequency analysis and exploratory scaling analysis, it turned out that these differentiations were not needed, i.e., they did not substantially contribute to the improvement of item fit statistics. As a consequence, partial-credit items were recoded to dichotomous items, i.e., additional categories were collapsed. For this, two different strategies were applied depending on the frequency distribution of each item: Either full credit (1) was given for all responses fulfilling at least one criterion or, in case this led to a better discrimination index and frequency distribution, full credit (1) was given for all responses that met two or more criteria. For example, in case of item example 3 (Supplementary Material, Fig. 5), a teacher's response was given full credit (score 1) only when she or he had provided at least two different techniques (e.g., using an acoustic signal, calling on an individual student by his name). All open-response items measuring classroom management expertise were coded on the basis of the coding manual. 40 questionnaires (about 20 % of all questionnaires) were randomly selected and coded by two raters independently of one another. Average of internal consistency was good ($M_{Kappa} = 0.81$, $SD_{Kappa} = 0.19$; cf. Fleiss and Cohen 1973).

4.3.3 Student rating of instructional quality

Specific measures of instructional quality related to classroom management were captured via student ratings. Although their validity is limited (e.g., they are suitable to assess classroom routines as part of a teacher's classroom management rather than to assess a teacher's didactic conceptions of teaching, cf. Baumert et al. 2004), we use such ratings, not least they provide an efficient way to capture instructional quality. Due to data collection constraints, this was only possible for our sample (2), i.e., second phase pre-service teachers. The students they taught were surveyed and thus asked to rate specific dimensions related to classroom management and learning support providing a multidimensional understanding of classroom management, namely a differentiation in organizational and instructional aspects (Gilberts and Lingnugaris-Kraft 1997). Organizational aspects were measured using the scale 'withitness' (4 items, e.g., 'Our teacher always knows exactly what happens in the classroom') and 'clarity of rules' (3 items, e.g., 'In the lesson it is clear what students are allowed to do and what they are not allowed to do'). Instructional aspects were measured using the scale 'clarity of teacher explanation' (3 items, e.g., 'Our teacher explains things step by step') and 'support' (4 items, e.g., 'The teacher additionally supports us when we need help'). Items for these scales were derived from the literature (e.g., Ramm et al. 2006). Items are Likert-scales with four categories ranging from 'not true' (1) to 'true' (4). Confirmatory factor analysis (CFA) using the software Mplus (Muthén and Muthén 1998–2006) was carried out showing a good fit for a model that specifies each scale as a latent variable ($\chi^2 = 2.11$, $p \leq 0.001$; RMSEA = 0.05, SRMR = 0.04). Scales are reliable ($\alpha \geq 0.6$) and their intra-class-correlations ('withitness' 0.14, 'clarity of rules' 0.08, 'clarity of teacher explanation' 0.10, 'support' 0.12) indicate there is substantial variation across the school classes thus making it necessary to conduct multi-level analysis (Julian 2001).

5 Results

5.1 Findings on dimensionality and inter-correlation (H1a)

In order to test our first hypothesis (H1a), with which we assume the classroom management expertise measure is positively correlated (medium effect size) but not identical with general pedagogical knowledge as assessed by the TEDS-M test, we compare a scaling model in which only one latent variable is specified by all classroom management expertise and general pedagogical knowledge test items (one-dimensional model) against a model that specifies classroom management expertise and general pedagogical knowledge as two latent variables (two-dimensional model).

Rasch-scaling analyses were done with the scaling software *ConQuest* (Wu et al. 1997). First, overall scaling of the one-dimensional model shows the overall measure including all classroom management expertise and general pedagogical knowledge test items is reliable (0.86) and variance of the latent variable (Theta-variance) is sufficient (0.84). However, in the two-dimensional model, reliability (Theta-variance) for classroom management expertise is 0.73 (0.91) and for general pedagogical knowledge is 0.85 (1.22). The inter-correlation is 0.65 thus showing the two constructs are clearly related to each other but not identical. Deviance statistics provides evidence the two-dimensional model fits significantly better to the data than the one-dimensional model (see Supplementary Material, Table 1). To strengthen our hypothesis, we alternatively conducted a confirmatory factor analysis specifying all items as categorical variables using the software *Mplus* (Muthén and Muthén 1998–2006) which provides similar findings. In this analysis, the two-dimensional model shows a slightly better fit (χ^2/df = 1.42, $p \leq 0.001$; RMSEA = 0.05) than the one-dimensional model (χ^2/df = 1.50, $p \leq 0.001$; RMSEA = 0.05). Latent inter-correlation between classroom management expertise and general pedagogical knowledge is 0.59 and thus even slightly lower than in the Rasch-scaling using the software *ConQuest* (Wu et al. 1997).

5.2 Inter-correlations of content dimensions (H1b)

Building on the two-dimensional model as analyzed in the previous Sect. 5.1 and using the software *Mplus* (Muthén and Muthén 1998–2006) again, general pedagogical knowledge was differentiated into four content dimensions (see Supplementary Material, Figure 2). This confirmatory factor analysis with four latent variables measuring general pedagogical knowledge and another latent variable measuring classroom management expertise shows a good fit (χ^2/df = 1.24, $p \leq 0.05$; RMSEA = 0.04). All latent variables are positively inter-correlated (see Supplementary Material, Table 2). However, as expected (H1b), the classroom management expertise measure is more highly inter-correlated with the general pedagogical knowledge content dimension 'classroom management/motivation' (0.68) than with the other general pedagogical knowledge content dimensions 'adaptivity' (0.45; z = 5.06, $p \leq 0.001$), 'structure' (0.59; z = 2.44, $p \leq 0.05$), and 'assessment' (0.24; z = 6.21, $p \leq 0.001$).[4]

5.3 Inter-correlations of cognitive demands (H1c)

In analogy to the analysis outlined in the previous Sect. 5.2, inter-correlations of cognitive demands were examined. The cognitive demands of the classroom management expertise measure 'accuracy of perception' and 'holistic perception' were combined into one dimension thus having a model with five latent variables only (see Supplementary Material, Table 3).[5] Model fit is good (χ^2/df = 1.28, $p \leq 0.05$; RMSEA = 0.04). As expected (H1c), 'interpretation/justification of action' as a cognitive demand of teachers' classroom management expertise correlates more highly with the general pedagogical knowledge cognitive demand 'generate' (0.70) than with the other general pedagogical knowledge cognitive demands 'recall' (0.57; z = −2,57, $p \leq 0.01$) and 'understand/analyze' (0.53; z = −3,74, $p \leq 0.001$). Similarly, 'perception' as a cognitive demand of classroom management expertise correlates more highly with 'generate' (0.60) than with 'understand/analyze' (0.46; z = −2,79, $p \leq 0.01$) and 'recall' (0.51; z = −1.59, p = 0.11), although the latter difference is not statistically significant.

[4] To compare differences in height of correlations, the significance test proposed by Meng, Rosenthal, and Rubin (1992) was applied.

[5] A model with six dimensions differentiating cognitive demands of the classroom management expertise in 'accuracy of perception', 'holistic perception', and 'interpretation/justification of action' did not converge due to the extremely high inter-correlation between 'accuracy of perception' and 'holistic perception'.

5.4 Expert-novice-comparisons (H2a, H2b)

To test the hypotheses related to our second research question (H2a, H2b), we again applied the two-dimensional model outlined in Sect. 5.1, but in addition to that we introduced expertise group assignment as two dichotomous predictors (see Supplementary Material, Figure 8). The model fit of this path model is good (χ^2/df = 1.42, $p \leq 0.05$; RMSEA = 0.05). In this analysis, 9.3 % of the classroom management expertise variance and 20.1 % of the general pedagogical knowledge variance is explained by the two predictors. In addition to this, Table 4 in the Supplementary Material provides means and standard deviations of test scores for the three groups of teacher expertise.

As expected (H2a), in-service teachers show better results than pre-service teachers (classroom management expertise: $\beta = 0.30$, $p \leq 0.001$; general pedagogical knowledge: $\beta = 0.42$, $p \leq 0.001$). Similar to this and as expected (H2b) pre-service teachers of the second phase of initial teacher education outperform pre-service teachers without any systematic teaching experience. However, the difference between the two groups of pre-service teachers is larger in their general pedagogical knowledge ($\beta = 0.30$, $p \leq 0.001$) than in their classroom management expertise ($\beta = 0.15$, $p \leq 0.10$).

5.5 Findings on the predictive validity (H3a, H3b)

The examination of our third research question requires us to conduct multi-level analysis, since the intra-class-correlations of the four student rating scales measuring instructional quality is substantial (between 0.08 and 0.14; see Sect. 4.4). Multi-level modeling was conducted again using the software *Mplus* (Muthén and Muthén 1998–2006). Due to small sample size of pre-service teachers on level 2 (n = 21) we use manifest variables and introduce only one predictor on level 2 per model.

As the findings indicate (see Supplementary Material, Table 5), organizational aspects of classroom management ('withitness' and 'clarity of rules') can be predicted by pre-service teachers' classroom management expertise and general pedagogical knowledge. However, classroom management expertise as a predictor is statistically significant (M1:

$\beta = 0.47$, $p \leq 0.01$; M3: $\beta = 0.36$, $p \leq 0.05$), whereas general pedagogical knowledge as a predictor is weaker (M2: $\beta = 0.32$, $p \leq 0.10$; M4: $\beta = 0.15$, n.s.), especially when explaining the instructional quality aspect of 'clarity of rules'. Classroom management expertise and general pedagogical knowledge are positively correlated ($0.20 \geq \beta \geq 0.28$) with instructional aspects of classroom management ('clarity of teacher explanations' and 'support'), but predictors are not statistically significant (see Supplementary Material, models M5 to M8 in Table 5). As hypothesized (H3a), there are positive correlations between classroom management expertise and specific measures of instructional quality. Moreover, as expected (H3b), classroom management expertise shows higher correlations with instructional quality aspects than general pedagogical knowledge.

6 Summary and discussion

Our study forwarded methodological consideration regarding the measurement of teachers' classroom management expertise (classroom management expertise), supporting the implementation of a novel video-based assessment approach that goes beyond the classical format of paper-and-pencil tests as hitherto provided by a study like TEDS-M (König et al. 2011). From a theoretical perspective and in relation to the requirements of classroom management, typical situations of classroom management and the knowledge-based processing of perceiving and interpreting classroom instruction were conceptualized. Classroom management expertise was empirically investigated by administering a test instrument that consists of four video clips used as item prompts and followed by test items related to these video clips. This study investigated the relationship between teachers' general pedagogical knowledge and classroom management expertise thus aiming to answer research questions (1) regarding the structural relation between classroom management expertise and general pedagogical knowledge, (2) regarding expert-novice differences in classroom management expertise and general pedagogical knowledge, and (3) regarding the predictive validity of classroom management expertise and general pedagogical knowledge.

6.1 Structural relation between classroom management expertise and general pedagogical knowledge

Regarding our first research question, findings from Rasch-scaling analyses (as well as from the confirmatory factor analysis approach) show teachers' general pedagogical knowledge and classroom management expertise are two different constructs, although they are substantially and positively inter-correlated (H1a). This is what research in other domains has also shown. For example, Kersting (2008, p. 857) reports a statistically significant manifest correlation ($r = 0.53$) between a paper-and-pencil test measuring mathematical content knowledge for teaching and a video-analysis instrument to measure teacher knowledge of teaching mathematics. The height of the latent correlation we found for pedagogy is similar to that reported by Kersting (2008) for the domain of mathematics. So, there seems to be a kind of analogy between the two very different assessments, supporting the construct validity of our approach. However, to go beyond a simple bivariate correlation of two constructs, we also made use of the differentiation of general pedagogical knowledge into subscales following previous work on general pedagogical knowledge which allowed us to localize the classroom management expertise measure in the broader field of teachers' general pedagogical knowledge. When differentiating general pedagogical knowledge into subscales, findings show that as expected (H1b) classroom management expertise is more highly inter-correlated with the content dimension of 'classroom management/motivation' than with any other content dimension of the general pedagogical knowledge test and, also as expected (H1c), more highly inter-correlated with the cognitive demand of 'generate' than with any other cognitive demand of the general pedagogical knowledge test. These findings assure construct validity of the novel video-based measure capturing classroom management expertise (Borsboom et al. 2004). They help us to understand that the conceptualization of classroom management expertise fits well into the framework of general pedagogical knowledge as previously developed in TEDS-M.

6.2 Expert-novice differences in classroom management expertise and general pedagogical knowledge

With our second research question, we followed the teacher expertise research paradigm (e.g., Berliner 1992, 2001; Bromme 1992, 2001; Sabers et al. 1991) and assumed differences between 'novice teachers' (i.e., student teachers at the start of initial teacher education), 'advanced beginners' (i.e., teacher candidates during their practical training at the end of initial teacher education), and 'expert teachers' (i.e., in-service teachers with, by average, 18 years teaching experience). Since we hypothesized such differences would mirror the idea classroom management expertise is developed during initial teacher education and in the course of professional teaching, we expected in-service teachers would outperform pre-service teachers in the classroom management expertise-test and general pedagogical knowledge-test (H2a) and pre-service teachers of the second phase of initial teacher education would outperform pre-service teachers without any systematic teaching experience (H2b). Evidence was found for both hypotheses with regards to general pedagogical knowledge, but only H2a could be confirmed with regards to classroom management expertise. So it seems that classroom management expertise, compared with general pedagogical knowledge, is much more dependent on the expertise level acquired during professional development, whereas general pedagogical knowledge can be acquired as early as during the theoretical initial teacher education at university (König 2013). This might be due to the specific situational nature of classroom management expertise when compared with general pedagogical knowledge. Against the findings on expert-novice differences, however, we consider our classroom management expertise measure an important instrument that could be applied in various research contexts, e.g., as an educational outcome in teacher education effectiveness research and research on the effectiveness of teacher professional development.

6.3 Predictive validity of classroom management expertise and general pedagogical knowledge

With our third research question we examined whether classroom management expertise, compared with general pedagogical knowledge, is a more proximal predictor of instructional quality. Due to data collection constraints this analysis was restricted to 21 pre-service teachers only, whose students had rated the instructional quality provided by these pre-service teachers. Findings from multi-level analysis show teachers' classroom management expertise can clearly be regarded as a predictor of aspects related to the organizational issues of classroom management ('withitness' and 'clarity of rules'). Moreover, the classroom management expertise turned out to be a more proximal measure than general pedagogical knowledge when explaining differences in instructional quality measures between school classes. Taking into account adequate mastery of classroom management is clearly related to student achievement (e.g., Hattie 2012) our finding that shows teachers' classroom management expertise predicts what actually happens in the classroom (captured by student ratings) has significant practical implications. This is especially true for the situation in Germany, where classroom management is not only an important facet of teaching quality, but the challenges for teachers to deal with classroom management issues such as disruptive behavior of students has increased during the last decade as, for example, the PISA cycles have shown (OECD 2013). Pre-service and in-service teachers thus should be provided with appropriate (e.g., practical) opportunities to learn in order to prepare them for classroom management challenges, including content that has been used to conceptualize classroom management expertise in our study.

6.4 Limitations and conclusion

Although findings are promising, at least three central limitations of our study have to be mentioned. First, we applied our novel approach to a sample of 188 pre-service and in-service teachers only (mainly due to data collection constraints). Therefore, replication studies using larger samples would be necessary to strengthen our work. This is especially true when taking into account that the examination of our third research question is based on a rather small teacher sample ($n = 21$). Second, although our in-service teacher sample contains the whole teaching staff of two schools with, by average, 18 years teaching experience, it is somehow problematic to simply denote them as 'experts' (cf. Berliner 2001). Future studies should look for another method to select in-service teachers (e.g., teachers denominated as experts by school principals) in order to clarify to what extent they may outperform pre-service teachers or even other groups of in-service teachers (e.g., early career teachers). And thirdly, instructional quality was measured using student ratings only, which in future studies could be, due to limited validity, extended by other methodological approaches such as video ratings.

To conclude, our findings let us assume a specific video-based assessment of classroom management expertise actually is of additional value when compared with a classical paper-and-pencil general pedagogical knowledge test. Evidence could be provided that classroom management expertise is of different quality compared with general pedagogical knowledge—it specifically can be characterized as being more of a procedural nature—, but at the same time it can be conceptually linked to general pedagogical knowledge. Thus our classroom management expertise measure can be regarded as a specific facet of general pedagogical knowledge that gives a more detailed insight into teacher professional knowledge and classroom management. It is, compared with general pedagogical knowledge, closer to what actually happens in the classroom, i.e., it is more specific and closer to the act of teaching than general pedagogical knowledge. One reason for this is that presumably video-cued testing may allow a more valid way of testing teacher competence (Blömeke et al. 2015; König 2015a, b). Seen from a different perspective, classroom management expertise, however, is not an isolated measurement instrument, but it can be connected to important research on general pedagogical knowledge theoretically and empirically: We have covered cognitive demands of perception and interpretation by applying a novel video-based assessment, we have related this to general pedagogical knowledge for an analysis of teacher competence, and we have related this to aspects of instructional quality in order to analyze indicators of teacher performance. Finally, we now have learned more about the "missing link between competence and performance" as focused on by this special issue, at least in the field of teacher professional knowledge and classroom management.

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.

Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W. & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In PISA-Konsortium (Eds.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland—Ergebnisse des zweiten internationalen Vergleichs* (pp. 314–354). Münster: Waxmann.

Berliner, D.C. (1992). The nature of expertise in teaching. In F.K. Oser, A. Dick, J.-L. Patry (Eds.), *Effective and Responsible Teaching* (Chap. 15, pp. 227–248). Jossey-Bass, San Francisco, CA.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*(5), 463–482.

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: a review of the state of research. *ZDM—The International Journal on Mathematics Education, 44*(3), 223–247.

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (2015). Beyond dichotomies: competence viewed as a continuum. *Zeitschrift für Psychologie, 223*, 3–13.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061–1071.

Bromme, R. (1992). *Der Lehrer als Experte: zur Psychologie des professionellen Wissens*. Bern: Huber.

Bromme, R. (2001). Teacher expertise. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 15459–15465). Amsterdam: Elsevier.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81.

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.

Doyle, W. (1985). Recent Research on Classroom Management: implications for Teacher Preparation. *Journal of Teacher Education, 36*, 31–35.

Doyle, W. (2006). Ecological Approaches to Classroom Management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 97–125). Mahwah: Erlbaum.

Evertson, C. M., & Weinstein, C. S. (Eds.). (2006). *Handbook of Classroom Management: Research, practice, and contemporary Research*. Mahwah: Erlbaum.

Fleiss, J. L., & Cohen, J. (1973). The Equivalence of weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement, 33*, 613–619.

Gilberts, G. H., & Lignugaris-Kraft, B. (1997). Classroom management and instruction competencies for preparing elementary and special education teachers. *Teaching and Teacher Education, 13*, 597–610.

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.

Good, T. L., & Brophy, J. E. (2007). *Looking in Classrooms* (10th ed.). Boston: Pearson Education.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Hawk, P. P., & Schmidt, M. W. (1989). Teacher Preparation. A Comparison of Traditional and Alternative Programs. *Journal of Teacher Education, 40*(5), 53–58.

Hill, H. C., Blunk, M., Charalambous, C., Lewis, J., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical Knowledge for Teaching and the Mathematical Quality of Instruction: an Exploratory Study. *Cognition and Instruction, 26*, 430–511.

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling, 8*(3), 325–352.

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*, 1–19.

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.

König, J. (2013). First comes the theory, then the practice? On the acquisition of general pedagogical knowledge during initial teacher education. *International Journal of Science and Mathematics Education, 11*(4), 999–1028.

König, J. (2014). *Designing an International Instrument to Assess Teachers' General Pedagogical Knowledge (GPK): Review of Studies, Considerations, and Recommendations*. Paris: OECD.

König, J. (2015a). Kontextualisierte Erfassung von Lehrerkompetenzen—Einführung in den Thementeil. *Zeitschrift für Pädagogik, 61*(3), 305–309.

König, J. (2015b). Measuring Classroom Management Expertise (CME) of Teachers: A video-based assessment approach and statistical results. Cogent Education, 2(1).

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., & Blömeke, S. (2013). Preparing teachers of mathematics in Germany. In J. Schwille, L. Ingvarson, & R. Holdgreve-Resendez (Eds.), *TEDS-M Encyclopaedia* (pp. 100–115). Amsterdam: IEA.

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education, 38*, 76–88.

König, J., Blömeke, S., Paine, L., Schmidt, B., & Hsieh, F.-J. (2011). General Pedagogical Knowledge of Future Middle School Teachers. On the Complex Ecology of Teacher Education in the United States, Germany, and Taiwan. *Journal of Teacher Education, 62*(2), 188–201.

König, J., & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand, 5*(1), 3–29.

Kounin, J. S. (1970). *Discipline and group management in classrooms*. Oxford: Holt, Rinehart & Winston.

Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*, 172–175.

Muthén, B. O., & Muthén, L. K. (1998–2006). Mplus (Version 4.2) [Computer software]. Los Angeles, CA.

OECD. (2013). *PISA 2012 Results: What Makes a School Successful (Volume IV): Resources, Policies and Practices*. Paris: OECD.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning. *Educational Researcher, 29*, 4–15.

Ramm, G., et al. (2006). *PISA 2003: Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.

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.

Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer"—Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. *Zeitschrift für Pädagogik, 56*, Beiheft, 296–306.

Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training, 2*(1), 43–65.

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

Slavin, R. E. (1994). Quality, appropiateness, incentive, and time: a model of instructional effectiveness. *International Journal of Educational Research, 21*, 141–157.

Swartz, C. W., White, K. P., Stuck, G. B., & Patterson, T. (1990). The factorial structure of the North Carolina teaching performance appraisal instrument. *Educational and Psychological Measurement, 50*(1), 175–182.

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Rowley, G., Peck, R., Bankov, K., Rodriguez, M., & Reckase, M. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries. Findings from the IEA teacher education and development study in mathematics (TEDS-M)*.

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.

Weinert, F. E. (2001). Concept of Competence: A Conceptual Clarification. In D. S. Rychenj & L.H. Salganik (ed.), *Defining and Selecting Key Competencies* (pp. 45–66). Göttingen.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: an examination of research on contemporary professional development. *Review of Research in Education, 24*, 173–210.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-Aspect Test Software [computer program]*. Camberwell: Australian Council for Educational Research.

Supplementary Material

Teacher professional knowledge and classroom management: On the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME)

Figure 1: Dimensions and topics covered in the TEDS-M test of GPK

Test dimensions	Topics covered by the test items
structure	- structuring of learning objectives - lesson planning and structuring the lesson process - lesson evaluation
classroom management/ motivation	- 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

Figure 2: Test design matrix

	recall	understand/ analyze	generate
structure			
classroom management/ motivation			
adaptivity			
assessment			

Figure 3: Item example 1 for assessing general pedagogical knowledge about "classroom management" as a topic and "understand/analyze" as a cognitive demand

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.	☐
B.	Yes, his behavior is caused by external intervention.	☐
C.	No, his behavior is caused by classical conditioning.	☐
D.	No, this behavior occurs spontaneously and is not learned.	☐

Figure 4: Item example 2 for assessing general pedagogical knowledge about "structure" as a topic and "generate" as a cognitive demand

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.

Figure 5: Item example 3 for assessing CME about 'accuracy of perception' as a cognitive demand

Please name four different techniques employed by the teacher to gain her students' attention.

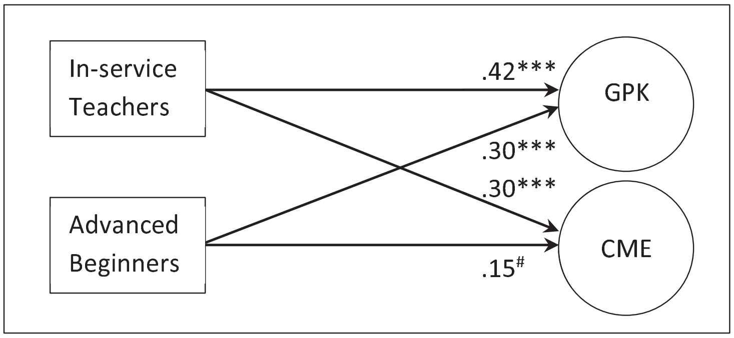
Figure 6: Item example 4 for assessing CME about 'holistic perception' as a cognitive demand

When does the situation displayed in the video take place?		
Please tick only one box:		
A.	At the beginning of the lesson (e.g., during the first 5 minutes).	□
B.	During the first third of the lesson	□
C.	During the last third of the lesson	□
D.	At the end of the lesson (e.g., during the final 5 minutes).	□

Figure 7: Item example 5 for assessing CME about 'interpretation/justification of action' as a cognitive demand

Which function does the seating arrangement of students have for the lesson shown?

Figure 8: Predicting GPK and CME by expertise group



*** p ≤ .001 ** p ≤ .01 * p ≤ .05 # p ≤ .10

Table 1: Deviance statistics from the one- and two-dimensional Rasch-scaling

model	deviance	number of estimated parameters	difference		
			deviance	number of parameters	*p*
one-dimensional (M1)	11,354.837	59	132.375	2	< .001
two-dimensional (M2)	11,222.462	61			

Table 2: Inter-correlations between GPK content dimensions and CME

			(1)	(2)	(3)	(4)
CME	(1)	classroom management				
GPK	(2)	adaptivity	.45***			
	(3)	structure	.59***	.54***		
	(4)	classroom management/ motivation	.68***	.69***	.77***	
	(5)	assessment	.24**	.38***	.33***	.31**

*** p ≤ .001 ** p ≤ .01 * p ≤ .05

Table 3: Inter-correlations between GPK cognitive demands and CME

			(1)	(2)	(3)	(4)
CME	(1)	perception (accuracy of perception/ holistic perception)				
	(2)	interpretation / justification of action	.97***			
GPK	(3)	recall	.51***	.57***		
	(4)	understand/ analyze	.46***	.53***	.45***	
	(5)	'generate'	.60***	.70***	.52***	.65***

*** p ≤ .001 ** p ≤ .01 * p ≤ .05

Table 4: Means and standard deviations of test scores for the three expertise groups

	Novices		Advanced Beginners		In-service Teachers	
	M	SD	M	SD	M	SD
GPK	19.9	6.6	25.9	4.1	28.6	4.2
CME	13.0	4.4	14.5	3.6	16.6	3.3

Table 5: Findings on the predictive validity of CME and GPK from multi-level analysis

	Withitness		Clarity of Rules		Clarity of Teacher explanations		Support	
	M1	M2	M3	M4	M5	M6	M7	M8
	β	β	β	β	β	β	β	β
CME	.47**	-	.36*		.20	-	.22	-
GPK	-	.32#	-	.15	-	.20	-	.28
R-square (between)	.22	.10	.13	.02	.05	.04	.05	.08

*** p ≤ .001 ** p ≤ .01 * p ≤ .05 # p ≤ .10