

*Profiles of mathematics teachers' competence and their relation to instructional quality*

**Sigrid Blömeke, Gabriele Kaiser,  
Johannes König & Armin Jentsch**

**ZDM**  
Mathematics Education

ISSN 1863-9690

ZDM Mathematics Education  
DOI 10.1007/s11858-020-01128-y



**Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.**



# Profiles of mathematics teachers' competence and their relation to instructional quality

Sigrid Blömeke<sup>1</sup> · Gabriele Kaiser<sup>2</sup> · Johannes König<sup>3</sup> · Armin Jentsch<sup>3</sup>

© The Author(s) 2020

## Abstract

Most of the research examining the relation between instructional quality (INQUA) and mathematics teachers' competence was restricted to a limited set of competence facets. Furthermore, mostly variable-oriented approaches were applied that assume homogeneity of teacher competence. In our study, teacher competence is conceptualized as a comprehensive multi-dimensional construct including a broad range of subject-specific and generic facets regarding mathematics teachers' knowledge, skills and beliefs, all assessed in a standardized way with well-established instruments. In our person-oriented approach, applying latent profile analysis, we examined whether it was possible to identify subgroups of teachers with different competence profiles. These profiles were then related to the generic and subject-specific INQUA implemented. The exploratory study was carried out with 77 secondary mathematics teachers from Germany as part of the projects TEDS Instruct and TEDS Validate. The data revealed four competence profiles that differed quantitatively and qualitatively. The profiles were related to different types of INQUA, and the relations indicate a need of strong levels of knowledge and skills for high INQUA but a lesser relevance of learning beliefs related to the dynamic nature of mathematics. In addition, our results indicate a need of a stronger subject-specific operationalization of INQUA in contrast to previously dominating generic conceptualizations.

**Keywords** Instructional quality · Teacher knowledge · Cognitive skills · Latent profiles · Mathematics teachers · Mathematical beliefs · Mathematics content knowledge

## 1 Introduction

The number of studies examining the relation between instructional quality (INQUA) implemented by mathematics teachers' and these teachers' competence has grown substantially over the course of the past decade (see e.g., Hill et al. 2008; Kaiser et al. 2015; Kunter et al. 2013). However, the state of research is still limited, amongst other aspects, due to methodological limitations. To the best of our knowledge, all studies published so far were based on a limited set of

measures. They either did not include a comprehensive set of generic and subject-specific measures of teacher competence, or the studies lacked standardized INQUA measures. Furthermore, most studies relied on the examination of linear (average) relationships between teacher competence and INQUA in a traditional variable-oriented approach, thus assuming homogeneity of competence profiles across teachers. However, it is not far-fetched to assume that different subgroups of teachers exist with qualitatively different competence profiles and that these subgroups may implement different types of INQUA.

The study presented in this paper, which belongs to follow-up-studies of the international 'Teacher Education and Development Study in Mathematics (TEDS-M)', intends to fill parts of these research gaps. To this end, we gathered a broad range of standardized data about lower-secondary mathematics teachers' competencies, observed several mathematics lessons of these teachers with a standardized observation protocol and applied an exploratory person-oriented analysis approach to these data. We assessed, firstly, mathematics teachers' subject-specific (i.e., mathematics

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11858-020-01128-y>) contains supplementary material, which is available to authorized users.

---

✉ Sigrid Blömeke  
sigrid.blomeke@cemo.uio.no

<sup>1</sup> University of Oslo, Oslo, Norway

<sup>2</sup> University of Hamburg, Hamburg, Germany

<sup>3</sup> University of Cologne, Cologne, Germany

content knowledge, MCK) and generic knowledge (general pedagogical knowledge, GPK). Secondly, we assessed teachers' subject-specific (i.e., speed in diagnosing student errors, MS<sub>Speed</sub>, and mathematics instruction related perception, interpretation and decision-making, M\_PID) as well as generic classroom-related skills (classroom management expertise, CME). Thirdly, we assessed these teachers' learning beliefs (related to the dynamic and static nature of mathematics as well as to the ability to learn mathematics).

Moreover, our study applied latent profile analysis (LPA) to this broad range of teacher measures. The profiles identified were then related to the instructional quality (INQUA) implemented in terms of the teachers' ability to manage the classroom, to support students, to activate students' cognition in a generic and subject-specific way as well as to implement mathematics teaching with high quality. The study was carried out with a small sample of 77 secondary mathematics teachers from four federal states in Germany and the two different types of schools existing there, an academic track and a non-academic track.

In the following, we present a focused literature review of the state-of-research on teacher competence (accomplished through the common variable-oriented approach) and its relation to INQUA. Then, we present our conceptual framework of teacher competence including the types of assessments used to cover the eight competence facets mentioned above. We also discuss the value-added elements provided by a person-oriented approach, before we present how the outcome (INQUA) was conceptualized and assessed in our study. Finally, we summarize the objectives for the present study in a set of hypotheses. Afterwards, the results of our study are delineated and discussed in a concluding section.

## 2 State of research

Weinert (2001) defined competence as a multi-dimensional construct functional for solving professional tasks and consisting of a broad range of cognitive abilities, motivation, volition and the social readiness to implement solutions in job-related contexts. Based on this definition, teacher-specific models of professional competence have been developed (e.g., by Baumert and Kunter 2006; for an overview see Blömeke and Kaiser 2017).

### 2.1 Teacher knowledge and beliefs

Teachers' *knowledge* was differentiated in content knowledge (in case of mathematics teachers this is MCK), pedagogical content knowledge (MPCK), and general pedagogical knowledge (GPK) (Shulman 1986). MCK includes knowledge about content domains such as number, algebra, geometry and data, which should provide teachers with the

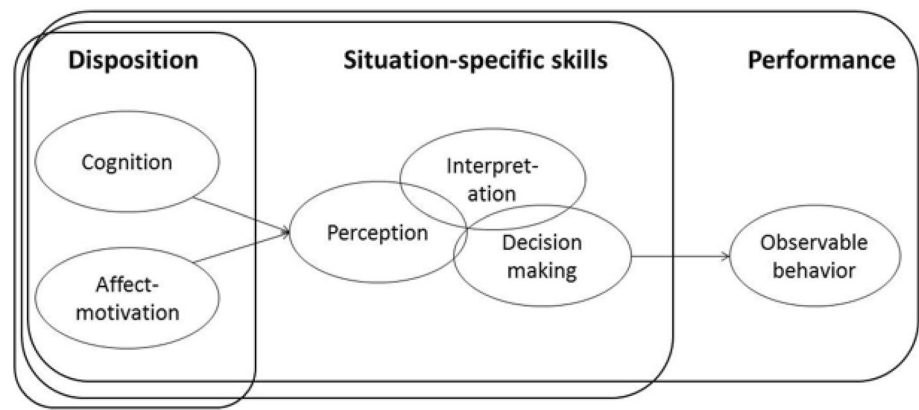
necessary background knowledge for teaching (e.g., Tatto et al. 2008, Kunter et al. 2013). MPCK covers curricular knowledge, knowledge of planning for mathematics teaching and learning of enacting mathematics (e.g., Tatto et al. 2008; Kunter et al. 2013). GPK includes knowledge about generic teacher tasks, not directly related to a specific subject, such as managing the classroom, motivating and supporting students, or assessing students (e.g., König et al. 2015). Empirical studies—for example COACTIV and TEDS-M and their various follow-up studies—provided evidence that these knowledge facets can be empirically separated but that MPCK strongly correlates both with MCK and GPK while these two are more distant from each other (e.g., Krauss et al. 2008, Baumert et al. 2010).

Kaiser and König (2019) provided a detailed summary of empirical results from this extensive research area. One result relevant to our study, which includes mathematics teachers from different types of schools, is the large difference in MCK depending on the opportunities to learn mathematics during teacher education. Teachers trained for academic school tracks or for a grade-span that includes upper-secondary grades achieved substantially better results in MCK than teachers trained for non-academic school tracks or a grade-span limited to lower-secondary or primary grades.

Besides cognitive facets, an understanding of competence based on the definition by Weinert (2001) includes affective-motivational characteristics. Note that this distinction serves mainly analytical purposes because affective-motivational characteristics often include a cognitive component too (Richardson 1996). Teacher *beliefs* are regarded as a crucial facet of affective-motivational characteristics (Blömeke and Kaiser 2017) because “beliefs might be thought of as lenses that affect one's views of some aspect of the world or as dispositions toward actions” (Philipp 2007, p. 259). There are many classifications of beliefs in the literature (Pajares 1992); especially important is a distinction of learning beliefs related to different views on the nature of mathematics ranging from viewing mathematics as a static, algorithm-driven body of facts and formulae best learned through memorizing, to a dynamic understanding of mathematics based on sense making and best learned through processes of enquiry (Goldin et al. 2016). Whether mathematical ability is regarded as innate, remaining fixed throughout a student's life, or whether it can be learned and thus further developed, has also played an important role in studies on teachers' beliefs (Tatto et al. 2008).

In the book edited by Blömeke et al. (2014), several chapters summarized what had been learned about mathematics teachers' beliefs from TEDS-M. German teachers supported significantly stronger learning beliefs related to the dynamic nature of mathematics than to its static nature. Regarding mathematics teachers from different types of schools, it was interesting to see that—compared to the large differences

**Fig. 1** Modeling teacher competence as a continuum (Blömeke et al. 2015a, p. 7)



in MCK—the differences in beliefs were much smaller or even non-existent between teachers trained for the academic school track or non-academic tracks.

## 2.2 Teachers' cognitive skills

This state of research was further developed by merging the previously separated understandings of teachers' professional competence from a situated perspective (Rowland and Ruthven 2011). The first perspective, reflected in the summary above and dominant over the last 2 decades, focuses on teachers' knowledge as traits rather stable across different classroom situations (e.g., Ball et al. 2008; Kunter et al. 2013). The second, more recent, perspective focuses on context-specific and situated aspects of teaching and learning (e.g., Kersting et al. 2012; Santagata and Guarino 2011). It departs theoretically from the concept of 'noticing', referring to teachers' skill in identifying what is noteworthy in a classroom situation (Sherin et al. 2011). One objective in this perspective is to identify competence facets more closely related to performance of teachers in the classroom (in terms of instructional quality; Charalambous and Praetorius 2018).

Blömeke et al. (2015a) integrated the cognitive and situated approaches by a new model on teacher competence as a continuum (Fig. 1). Competence and performance are distinguished in this model but assumed to be functionally related to each other, and competence is conceptualized as a multi-dimensional latent construct that manifests in teaching performance and includes all mental resources necessary to perform (Klieme et al. 2008). Despite substantial differences in construct conceptualization, labelling and operationalization, several empirical studies can be related to this approach of regarding teacher knowledge and skills as integrated facets of teacher competence underlying classroom performance in terms of instructional quality, which in turn is hypothesized to affect student achievement, in particular the studies by Bruckmaier et al. (2016), Kniewel et al. (2015), and Stürmer and Seidel (2015).

In the model formulated by Blömeke et al. (2015a), teacher dispositions represent the potential a teacher brings to the classroom. Dispositions include cognitive characteristics such as teachers' professional knowledge, but also characteristics with an affective-motivational component such as teachers' beliefs. This part of the model has similarities with models developed, for example, by Ball et al. (2008) and Baumert and Kunter (2006). Differences are limited to construct conceptualization, and operationalization within these dispositional facets.

In contrast to the latter two models, the Blömeke et al. (2015a) competence model also includes teachers' situation-specific cognitive skills. In this part, the model is more in line with research on noticing, e.g. by Kersting et al. (2012) or Santagata and Guarino (2011). Teachers' skills include the ability to perceive and interpret what is going on in the classroom and then to make decisions (PID). In more detail, taking into account Schön's (1983) concepts of reflection in and on action, teachers' situation-specific skills are conceptualized as a facet of their cognitions, not being equal to observable teaching behavior but representing cognitive processes prior to, during, and following performance in the classroom (Star and Strickland 2008). Diagnostic skills play an important part within these teaching and learning processes, especially a fast diagnosis of students' errors or misconceptions (Leuders et al. 2018).

The relation between the different cognitive facets of teacher competence was evaluated in several studies. Blömeke et al. (2016) provided evidence for a two-dimensional structure of mathematics teachers' competence, differentiating between one factor related to knowledge and another one related to situation-specific cognitive skills. Bruckmaier et al. (2016) found moderate positive correlations between MCK and MPCK and teachers' decision-making. The relation between GPK and noticing was evaluated by König et al. (2014), who found low to moderate relations between teachers' general pedagogical knowledge and their generic perception and interpretation skills. A comparative study across a Western and an East Asian context revealed

relative strengths of German teachers in noticing from a general pedagogical perspective (P\_PID), whereas Chinese teachers were relatively strong in noticing from a mathematics pedagogical perspective (M\_PID) (Yang et al. 2018).

Whereas these studies assumed that all teachers in a sample had the same competence profile, a study with the aim of identifying different competence profiles was carried out with TEDS-M data of mathematics teachers for primary schools (Blömeke et al. 2012). Subject-specific knowledge and beliefs were used in a latent profile analysis (LPA) where for each country profiles were identified that differed either quantitatively or qualitatively. Quantitative differences were found, for example, in Germany, Norway or the US where a profile existed that combined higher MCK with higher MPCK, as well as stronger dynamically—but weaker statically—oriented learning beliefs related to the nature of mathematics. At the same time another profile existed within these countries, where the opposite applied. A profile that differed qualitatively was, for example, found in Russia where higher MCK and MPCK was combined with stronger learning beliefs related to the static nature of mathematics.

### 2.3 Teacher performance

Instructional quality (INQUA) reflects observable teaching performance in terms of instructional processes implemented in classrooms. Several frameworks and instruments exist that assess either generic (e.g., Fauth et al. 2014) or subject-specific facets of instructional quality (often in mathematics, e.g., Hill et al. 2008; more recently also in science, e.g., Carlson and Daehler 2019; see an overview by Schlesinger and Jentsch 2016). Recently, a combination of these perspectives has been requested (e.g., Lipowsky and Bleck 2019).

Generic approaches usually comprise three facets, in particular in the German context (Praetorius et al. 2015), namely, classroom management, student support, and cognitive activation, sometimes named slightly differently. Classroom management means an efficient use of time and the prevention of or dealing with disorder in the classroom (Evertson and Weinstein 2013). The second INQUA facet, student support, comprises either addressing individual learning needs of students or fostering a prolific classroom climate (Fauth et al. 2014, Gräsel et al. 2017). Our study follows the first conceptualization. Cognitive activation refers to whether teachers' instruction is cognitively challenging for students (Klieme et al. 2009, Lipowsky et al., 2009). Cognitive activation was originally conceptualized as a generic facet without referring to a specific subject (Klieme and Rakoczy 2008), but has been identified as particularly relevant for subject-specific INQUA in recent studies (Praetorius et al. 2015).

Subject-specific INQUA represents mathematics-related quality. In particular, this aspect relates to how teachers address mathematical concepts and interact mathematically with students while using mathematical terms, explaining mathematical procedures, providing feedback or dealing with student errors (Hiebert and Grouws 2007; Schlesinger et al. 2018). Studies provided evidence on the predictive validity of INQUA on cognitive and affective-motivational student outcomes (e.g., Kunter et al. 2013; Praetorius et al. 2018).

Variable-oriented research revealed the relevance of teacher knowledge and skills for observable teaching performance in terms of INQUA. Kunter et al. (2013) found that lower-secondary mathematics teachers with higher MPCK implemented higher cognitive activation and student support (but not classroom management). Similarly, preschool teachers with more MPCK created opportunities for children to learn that were of higher quality (Lee et al. 2003; with respect to other domains see Gold et al. 2013; Stürmer et al. 2013). Bruckmaier et al. (2016) revealed a low but significant positive relation between mathematics teachers' decision-making skill and their cognitive activation (but not their student support or classroom management). GPK has also been found to be a significant positive predictor for classroom management and student support (Depaepe and König 2018; Voss et al. 2014), whereas the relation with cognitive activation varied.

The state of research on the relation between teachers' beliefs and INQUA is limited and mixed (Simmons et al. 1999). This result applies also to the German context. Whereas Voss et al. (2013) found positive associations between constructivist beliefs and instructional practices and Bruckmaier et al. (2016) found positive associations between constructivist beliefs and teachers' decision-making skills (not controlling for other competence facets), a study by Kunter et al. (2013) found that higher constructivist beliefs were not systematically associated with cognitive activation or student support, but were associated significantly with lower classroom management when controlling for MPCK.

## 3 Conceptual framework and research questions

In this study we regard competence as functionally related to performance, as displayed in Fig. 1. Our aim is to identify different profiles in how teachers' dispositions, consisting of their knowledge and affect-motivation, are related to situation-specific skills within teachers and how these profiles are associated with observable behavior in the classroom. Instructional quality is used as an indicator for the latter.

### 3.1 Teacher competence

Teachers' *cognitive* competence facets are conceptually differentiated into subject-specific (MCK, MPCK) as well as generic dispositions (GPK) on the one hand, and subject-specific (M\_PID, speed in diagnosing student errors MSpeed) as well as generic (classroom management expertise CME) perception, interpretation and decision-making skills on the other hand. Our study covered these cognitive facets as comprehensively as possible and included the five most distinct facets MCK, GPK, M\_PID, MSpeed and CME.<sup>1</sup>

The MCK and GPK assessments were computerized abbreviated versions of the TEDS-M tests and were validated through expert reviews and empirical studies (e.g., Blömeke et al. 2015b; König et al. 2014). MCK includes factual knowledge of mathematics and conceptual knowledge of its organizing principles, for example where a topic is placed in the universe of mathematics. The test covered number, algebra, and geometry and, to a lesser extent, data (Tatto et al. 2008). In addition, three cognitive dimensions were covered, namely knowing, applying, and reasoning. GPK includes preparing and structuring lessons, motivating student learning and managing the classroom, dealing with heterogeneous learning groups and assessing students. Similarly to assessment of MCK, three cognitive processes were distinguished, namely, recalling, understanding, and generating. For MCK and GPK item examples see the Electronic Supplementary Material.

M\_PID was conceptualized as explained in Sect. 2.2. Video- and computer-based assessments were used to address the contextual nature and complexity of classroom situations. Three scripted video-vignettes of 4-min length served as substitutes for genuine situations. They displayed mathematical topics typically taught in German schools in grades 8–10 in different phases of teaching, such as the introduction of a mathematical task, or student work on a task followed by a whole-class discussion of the results. The assessment was validated in several studies (e.g., Hoth et al. 2016). These studies have, among other aspects, provided evidence that testlet effects related to the videos were negligible (Blömeke et al. 2015b).

Perceiving and adequately responding to student errors is part of mathematics teachers' daily activities in the classroom (Pankow et al. 2018) and has been stressed in the context of adaptive teaching (Südkamp and Praetorius 2017).

<sup>1</sup> As pointed out above, MPCK overlaps conceptually and empirically with MCK and GPK (Depaepe et al. 2013). In our study, MPCK correlated particularly highly with MCK ( $r=0.80$ ). In an LPA (for details see Sect. 3.2), such correlations can create residual covariances and thus distort the number of profiles identified. We therefore excluded MPCK.

Heinrichs and Kaiser (2018) have developed a model for the perception, interpretation and decision-making regarding errors in mathematics education according to which the perception of student errors is the first step of a cyclic process leading to the development of hypotheses about the causes of these errors, and the making of a decision about the instructional approach appropriate for dealing with the error (see also Hoth et al. 2016). The time mathematics teachers have for noticing student errors is limited though. A time-limited computer-based test displaying typical student errors was applied in the present study to assess the fast identification of student errors.

As a generic skill measure, we used the classroom management expertise (CME) video-based assessment (König et al. 2015) that consists of four vignettes referring to typical classroom management situations in which teachers have to manage transitions, instructional time, student behavior, and instructional feedback. A variety of classroom contexts regarding school grades, subjects, and the composition of the learning group are represented (Kounin 1970). The items require accurate and holistic perception as well as interpretation. Validation studies revealed that the CME score represents an ability only slightly influenced by testlet effects related to the four video vignettes (König et al. 2015).

A well-established set of learning beliefs was used in this study, namely those related to the dynamic or static nature of mathematics (Tatto et al. 2012). In a dynamic view, learning mathematics is seen as a process of enquiry and the application-related character of mathematics is emphasized. In a static view, respondents tend to see mathematics as a set of procedures to be learned with strict rules as to what is correct and what is incorrect.<sup>2</sup>

Teachers' beliefs about mathematics ability reflect the idea that students are born with such an ability and that this ability is fixed throughout their lives. When holding such a belief, teachers may show different expectations towards children, viewed as more or less able (Hart and Drummond 2014). This belief has also been studied in the context of TEDS-M.

### 3.2 Teacher profiles

Traditional variable-oriented approaches assume sample homogeneity with respect to relationships between variables.

<sup>2</sup> As an alternative we could have used teachers' constructivist and transmission-oriented beliefs. However, we decided to include this set of beliefs that was more directly related to the learning of mathematics (see the item examples in the Electronic Supplementary Material) and had better measurement properties given that they were used as manifest variables. Due to limited sample-size and multi-collinearity challenges, it was not possible to include both sets of beliefs at the same time.

However, given that the eight facets of teacher competence included in this study reflect a high degree of complexity, subgroups of teachers may exist with distinct competence profiles that can be characterized as stronger on a subset of competence facets and weaker on another. Given that the interplay of competence facets is assumed to be of particular relevance for performance (Schoenfeld 2010), it is important to identify these profiles so that a teacher can be characterized as more or less competent (Oser 2013).

### 3.3 Instructional quality and its relation to teacher competence

Given that it may be possible to identify different profiles of mathematics teachers, Blömeke and Kaiser (2017) raised the questions of which level of dispositions and skills is enough for a teacher to be called ‘competent’, and whether the different facets of competence can compensate for each other (Koeppen et al. 2008). Does a teacher need to have all types of dispositions and skills to be able to perform successfully in class or is it sufficient, for example, to have strong mathematics-related skills? To obtain a first answer to these questions, an outcome variable was needed that indicated different levels of performance quality, in our case instructional quality. The observation protocol used to assess INQUA was validated in several contexts (e.g., Schlesinger et al. 2018).

### 3.4 Research objectives and hypotheses

As a first step, we wanted to identify competence profiles of mathematics teachers by applying LPA to our data. We hypothesized (**H1**) that several distinct groups of mathematics teachers existed, which revealed either quantitative differences in competence (e.g., high, medium, and low knowledge combined with corresponding levels of skills and more or less favorable types of beliefs) or qualitative differences (e.g., one group with high MCK but low CME vs. a group with low MCK but high CME).

Second, the relationship of these competence profiles to INQUA was examined. We hypothesized (**H2**) differential relationships: in case of quantitative differences, we expected that a group with higher knowledge and skills and more dynamically- or less statically-oriented learning beliefs would implement higher levels of INQUA on all four facets. In case of qualitative differences, we expected that the level of INQUA would depend on the specific profile.

As a final step, the relationship of these profiles to school types was examined given that variable-oriented research had pointed to large differences in MCK but less pronounced differences in beliefs and levels of generic competence facets between mathematics teachers at the Gymnasium as the academic school track, on the one hand, and mathematics teachers at non-academic school types on

the other hand. We hypothesized (**H3**) that the proportion of teachers per school type may vary across profiles with higher proportions of teachers at a Gymnasium in the profile with higher MCK.

## 4 Methodology

### 4.1 Sample

In the context of two studies in Germany that examined the professional competence of practicing mathematics teachers, altogether 77 teachers took the full range of competence assessments discussed above. The studies had the same study design but took place in different regions and in different years, namely, first TEDS-Instruct in the state of Hamburg, and then subsequently TEDS-Validate in the states of Thuringia, Hesse and Saxony. Overall testing time was about 5 h. About 60% were teaching at a Gymnasium, the academic school track ( $n=47$ ), while the others were teaching at non-academic schools ( $n=30$ ). About half of the sample was recruited in Hamburg ( $n=38$ ) while the other mathematics teachers were from Thuringia ( $n=20$ ), Hesse ( $n=14$ ) or Saxony ( $n=5$ ). Average class size was 23 students.

### 4.2 Instruments

Teachers took all tests individually at home and received an honorarium to compensate for their efforts. They could pause test-taking once. For item examples see the Electronic Supplementary Material. Constructed responses were coded based on coding manuals. Scaled scores were created by applying classical test theory using SPSS (sum scores with Cronbach’s  $\alpha$  as the reliability parameter) and item-response theory using the software package Conquest (Wu et al. 1997; 1-dimensional Rasch scaling, Expected A Posteriori reliability measure EAP). Items omitted or not reached were treated as wrong responses during the scaling process. The full TEDS-Instruct and TEDS-Validate samples, also those teachers who did not take all assessments, were included in the scaling to increase parameter stability. All scales were transformed to a mean of 500 and a standard deviation of 100 to facilitate understanding.

MCK was assessed with 27 multiple-choice, complex multiple-choice, and constructed-response items. Reliability of the MCK scale was  $Cr \alpha=0.81$  or  $EAP=0.79$  respectively. The GPK test included five multiple-choice and ten open-response items. Scale reliability was  $EAP=0.90$  or  $\alpha=0.71$ .

M\_PID was assessed with 16 items using Likert scales, and 18 constructed-response items formed the M\_PID scale. Correct answers to the Likert-scale items were developed



during an expert meeting including experienced mathematics teachers and mathematics education researchers. Intercoder reliability varied between about  $\kappa_{\min} \approx 0.70$  and  $\kappa_{\max} = 1.00$  and was on average very good ( $\kappa_{\text{ave}} \approx 0.90$ ; Landis and Koch 1977).

The MSpeed test consisted of 16 items. All items contained three student's solution to a specific theme of secondary school mathematics, e.g. addition of fractions, in which one of the shown solutions was wrong. The test participants had 4 s time to recognize the wrong solution (Pankow et al. 2018). Reliability of the scale was  $EAP = 0.64$ .

CME assessment was video-based: the video assessment consists of 24 test items (5 multiple-choice and 19 constructed-response items). Scale reliability was  $EAP = 0.75$  or  $\alpha = 0.71$ .

Beliefs were evaluated with items using a 6-point Likert scales from 'do not agree at all' to 'fully agree'. Scale scores represented the mean across all items as long as at least two items had been rated. Reliability of the scales varied between  $\alpha = 0.64$  for dynamically-oriented learning beliefs and  $\alpha = 0.74$  for beliefs regarding mathematics ability as innate and fixed.

INQUA was evaluated with live ratings. The observation protocol consists of 18 items that had to be rated on four-point Likert scales ranging from 1 ('does not apply at all') to 4 ('does fully apply'). 10 raters with at least a Bachelor's degree in mathematics education received 30 h of theoretical and practical training. For each lesson two raters were randomly distributed to take the observation task. Inter-rater reliability and validity of the rating was supported by a rater manual that included typical classroom examples (Schlesinger et al. 2018). Each mathematics teacher was observed during four lessons at 20 min intervals, resulting in eight time points for each teacher. Inter-rater reliability was good ( $ICC > 0.80$ ). Scaling analysis summarized all information using average scores. Reliability of the scales was good ( $\alpha = 0.73\text{--}0.87$ ).

### 4.3 Data analysis

To examine our first hypothesis, we conducted latent profile analysis (LPA) using the software package Mplus 8.2 (Muthén and Muthén 1998–2018). LPA is a model-based exploratory method to classify similar objects—in our case mathematics teachers—into homogenous groups where the number of classes as well as their properties are unknown and inferred from the data (McLachlan and Peel 2000). LPA is particularly beneficial in complex analyses when a larger number of continuous and moderately correlated<sup>3</sup> variables

are included in an analysis, because the number of interactions that can be included in traditional multiple regression is limited.

Since LPA is an exploratory approach, it is not necessary to decide about the number of classes or the type of relationship between the variables beforehand. The results describe the conditional probabilities of having knowledge, beliefs and skills given the class membership, so that each class is characterized by a specific pattern, called 'profile'. Teachers are assigned to the class for which their observed response pattern is most probable. Since a person-oriented approach does not assume sample homogeneity, the objective of an LPA is to identify qualitatively different profiles, not only quantitative differences (e.g., higher estimates on all variables in class 1 versus lower estimates in class 2). A profile with such quantitative differences would be reflected in a variable-oriented approach too.

For model parameter estimation a robust maximum-likelihood estimator (Sass et al. 2014) and a sandwich-type covariance matrix were applied, in order to achieve precise estimations of standard errors (Satorra and Bentler 2001) and Chi square statistics robust to non-normality of the data (Yuan and Bentler 2000). Full Information Maximum Likelihood (FIML) estimation, integrating missing data analyses and parameter estimation under the missing at random assumption, was used to handle partially missing data (Little and Rubin, 2014). However, the proportion of missing data was low; covariance coverage was at least 95% in all cases.

Given the sample size, we used manifest scores in the LPA. We evaluated the classification quality based on an aggregated uncertainty measure called "entropy" (Ramawamy et al. 1993). In its rescaled version implemented in MPlus, it has an interval of [0, 1] with estimates close to 1 for well-separated classes and estimates close to 0 for ill-fitting models. In addition, the estimate of the mean probability for teachers' most likely class membership was taken into account. We also evaluated relative fit criteria (Nylund et al. 2007), where lower absolute values indicate a better-fitting model. The Akaike Information Criterion AIC and the adjusted Bayesian Information Criterion  $BIC_{\text{adj}}$  (Schwartz 1978) are measures of the goodness of fit of an LPA model that consider the number of parameters and the number of observations.

Finally, we applied two likelihood-ratio tests, the Lo–Mendell–Rubin adjusted LRT test (Lo et al. 2001) and the bootstrapped parametric likelihood ratio test (BT LRT; McLachlan and Peel 2000), which compare the current solution against one with one fewer class. A low (significant) *p* value indicates that the current model is more appropriate to describe the data than a  $k - 1$  class model. Since the BT LRT is regarded as most important, we relied heavily on this one (Nylund et al. 2007). As suggested in the literature, our final decision about the number of classes was also based on the

<sup>3</sup> LPA is thus slightly different to a latent class analysis (LCA), which is typically applied to categorical data and includes a local independence assumption (i.e. uncorrelated indicators) within classes.

**Table 1** Mean scores and correlations for the competence facets included in the LPA

	MCK	MSPEED	M_PID	GPK	CME	Dynamic	Static	Ability
MCK	507							
MSPEED	0.42*	489						
M_PID	0.33*	0.24*	507					
GPK	0.44*	0.35*	0.34*	485				
CME	0.32*	0.19	0.58*	0.29*	494			
Dynamic beliefs	-0.25*	-0.16	0.09	-0.18	0.02	534		
Static beliefs	-0.12	-0.17	-0.09	-0.06	0.17	-0.07	374	
Fixed ability	-0.01	0.06	-0.26*	0.00	-0.04	-0.11	0.34*	246

*MCK* mathematics content knowledge, *MSpeed* fast perception of student errors, *M\_PID* mathematics instruction related perception, interpretation and decision-making, *GPK* general pedagogical knowledge, *CME* classroom management expertise; mean values are displayed on the diagonal

\* $p < 0.10$

conceptual interpretability of the classes (Lubke and Muthén 2005). Given the sample size, we defined a threshold of 10% ( $n=8$ ) as desirable class size.

To examine our second and third hypotheses, the four INQUA facets and the school type were specified as outcomes. In Mplus, an estimation procedure for small sample sizes is implemented via the DU3STEP command where unequal means but equal variances of the outcome variables are assumed across classes (Asparouhov and Muthén 2013). The level of significance was set to  $p < 0.10$  due to the sample size.

## 5 Results

### 5.1 Descriptive results

The descriptive results for the eight competence facets are displayed in Table 1. Means are displayed on the diagonal whereas correlations can be found below that line. The means revealed that the 77 teachers included in this study had a slightly higher mean on the subject-specific competence facets MCK ( $M=507$ ) and M\_PID ( $M=507$ ) than the full sample, and at the same time a lower mean on the two generic competence facets GPK ( $M=485$ ) and CME ( $M=494$ ), but also on the subject-specific facet MSpeed ( $M=489$ ). Correlations between teachers' cognitive competence facets (knowledge and skills) were all positive and of moderate to medium size ( $r=0.19-0.58^*$ ).

Teachers in this sample agreed with learning beliefs related to a dynamic view on mathematics ( $M=534$ ) and rejected strongly learning beliefs related to a static view on mathematics ( $M=374$ ). They rejected even more strongly that mathematics ability remains fixed throughout the life and cannot be further developed ( $M=246$ ). The two beliefs facets on the nature of mathematics represented largely independent views ( $r=-0.07$ ) while the static

view on learning mathematics and the view that mathematics ability is fixed were significantly positively related to each other ( $r=0.34^*$ ).

Relationships between teachers' knowledge and skills and their beliefs ranged from moderately negatively to insignificant ( $r=-0.26^*$  to  $0.17$ ). As a tendency, the subject-specific competence facets MCK, MSpeed and M\_PID were negatively related to the three beliefs facets, which means that higher mathematics-related knowledge or skills was associated with lower agreement to a dynamically- or a statically-oriented learning belief or mathematics ability as innate and not learnable. The two generic facets GPK and CME did not show a systematic relationship to teachers' beliefs.

### 5.2 Number of classes

In an iterative process, we steadily increased the number of classes, until the fit indices indicated that a lower number of teacher profiles was sufficient and the added class not needed. The latter applied to the 5-class solution where the bootstrapped LRT pointed to a  $k-1$  class model (see Table 2). With the exception of the LMR LRT test that rejected any model with more than one class, also all other indicators pointed to a 4-class solution as the most appropriate. Since this solution also revealed the most interesting teacher profiles with qualitative differences, while the other solutions only contained quantitative differences, we decided to select the 4-class solution.

### 5.3 Competence profiles of mathematics teachers (H1)

The four groups of mathematics teachers differed quantitatively and qualitatively (see Table 3). Despite the large standard errors given the relatively small group sizes, the

**Table 2** Fit indices for different numbers of classes (competence profiles)

#Class	#Par	AIC	BIC <sub>adj</sub>	Entropy	Class prob.	Class size	LMR LRT	BT LRT
2	25	6988.7	6968.5	0.83	0.95/0.97	82%/18%	p=0.20	p<0.001
3	34	6978.3	6950.8	0.79	0.92/0.89/0.90	61%/23%/16%	p=0.74	p<0.05
4	43	6968.7	6933.9	0.83	0.90/0.90./96/0.96	35%/30%/25%/10%	p=0.28	p<0.001
5	52	6964.3	6922.3	0.84	0.90/0.89/0.89/0.95/0.96	35%/26%/21%/10%/8%	p=0.59	p=0.43

AIC Akaike Information Criterion, BIC<sub>adj</sub> adjusted Bayesian Information Criterion, LMR RT Lo–Mendell–Rubin adjusted LRT test, BT LRT bootstrapped parametric likelihood ratio test

**Table 3** Descriptive results for the four-class solution

	MCK	MSPEED	M_PID	GPK	CME	DYN	STAT	FIXED
1 (n=27)	593 (18)	571 (17)	556 (33)	531 (21)	526 (32)	515 (19)	355 (19)	252 (17)
2 (n=23)	474 (25)	435 (32)	453 (18)	454 (26)	487 (21)	540 (14)	412 (13)	313 (13)
3 (n=19)	473 (25)	448 (37)	553 (21)	479 (24)	541 (26)	552 (13)	366 (15)	179 (15)
4 (n=8)	407 (30)	478 (23)	367 (28)	424 (37)	273 (24)	538 (16)	345 (32)	202 (30)

MCK mathematics content knowledge, MSpeed fast perception of student errors, M\_PID mathematics-related perception, interpretation and decision-making, GPK general pedagogical knowledge, CME classroom management expertise, DYN dynamically-oriented learning beliefs about mathematics, STAT statically-oriented learning beliefs about mathematics, FIXED beliefs about mathematics ability as innate and fixed

mean differences in the eight competence facets were statistically significant with groups 1 and 4 as the extremes.

The first and largest teacher profile was characterized by high levels of knowledge and skills compared to the other three profiles. This applied particularly to subject-specific facets but to some extent also generically. These teachers differed cognitively from the mean of the full sample by one quarter to a full standard deviation. Teachers with such a profile had an only slightly dynamic view on learning mathematics while they strongly rejected learning beliefs related to the static nature of mathematics as well as the view that mathematics ability remains fixed throughout one's life.

Teacher profiles 2 and 3 were characterized through medium to slightly below medium levels of MCK, MSpeed and GPK as well as very dynamic beliefs about learning mathematics. In contrast to these similarities, these two profiles were at the same time characterized through substantially different levels of skills and the other two beliefs facets. The second teacher profile displayed medium levels of M\_PID and CME and rather neutral (compared to the other groups) learning beliefs related to the static nature of mathematics and a fixed nature of mathematics ability. The third teacher profile displayed high levels of M\_PID and CME and low agreement with the two types of beliefs.

The fourth teacher profile included a small group of mathematics teachers only. The profile was characterized by low levels of subject-specific and generic knowledge and skills. With the exception of MSpeed, the difference from the mean of the full sample was between one to two standard deviations, and all mean comparisons with the first teacher group

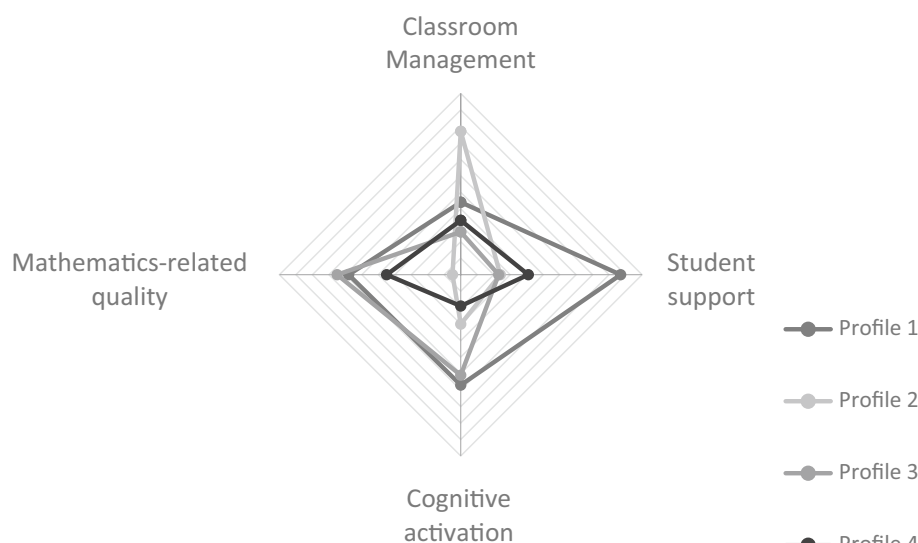
as well as in case of M\_PID and CME with the second and third groups were significantly different. However, similarly to the other three groups, also these teachers believed in inquiry learning related to the dynamic nature of mathematics and rejected strongly that mathematics ability is innate and fixed.

### 5.4 Relation between competence profiles and INQUA (H2)

The data supported our hypothesis that different profiles are associated with different types of instructional quality. Mathematics teachers belonging to the first profile—characterized through high levels of knowledge and skills in all respects and strong objections against a static view on learning mathematics or on mathematics ability remaining fixed throughout one's life—implemented high levels of student support (M = 54.7\*), mathematics-related quality (M = 51.8\*) and descriptively also of cognitive activation (M = 51.7; see Fig. 2).<sup>4</sup> In contrast, the fourth teacher profile—characterized by the lowest levels of knowledge and skills but strong beliefs—did not excel on any of the four INQUA facets.

<sup>4</sup> M = 50 for all INQUA scales; \*p < 0.10 based on equality tests of INQUA means across the four teacher profiles using the 3-step procedure implemented in Mplus. Classroom Management: p < 0.10 for profile 2 vs. profile 3, student support: p < 0.10 for profile 1 vs. profiles 2 and 3, mathematics-related quality: p < 0.10 for profiles 1 and 3 vs. 2.

**Fig. 2** Instructional quality by teacher profile



Teacher profiles 2 and 3, which were similar in many respects but differed with respect to their levels of cognitive skills, delivered different types of INQUA. Teachers belonging to the third profile, characterized by stronger cognitive skills, implemented a significantly higher level of mathematics-related quality ( $M = 52.5^*$ ) and descriptively also of cognitive activation ( $M = 51.1$ ), whereas the second teacher profile, characterized by a medium level of cognitive skills, did significantly better when it came to classroom management ( $M = 53.7^*$ ).

### 5.5 Relation between competence profiles and school type (H3)

Against our hypothesis, a significant difference between the four teacher profiles regarding their association with school types did not exist. However, on a descriptive level a tendency was visible, namely a steady decrease in the proportion of teachers in the academic track (Gymnasium) from 72% in class 1 to 50% in class 4.

## 6 Discussion and conclusions

Teacher competence has been conceptualized as a multi-dimensional construct in this study including subject-specific and generic facets of mathematics teachers' knowledge, skills and beliefs. Most of the research up to now was restricted to a limited set of competence facets and variable-oriented approaches that assumed homogeneity in teachers' competence profiles. In contrast, using our person-oriented approach, with data about eight competence facets from a small sample of mathematics teachers in Germany ( $n = 77$ ), we explored whether sub-groups exist.

The data supported the hypothesis that mathematics teachers may not be as homogeneous as was assumed in common variable-oriented approaches (e.g., linear regression). The profiles discovered not only differed quantitatively, in that sense that several competence facets were higher (or lower) in one teacher group than in another, but also qualitatively in that sense that similar levels of one set of competence facets appeared together with different levels of others.

The largest group of mathematics teachers (#1) was characterized by the highest levels of subject-specific and generic knowledge and skills whereas the smallest group performed the lowest on all these assessments. Cognitive skills made the biggest differences between the two other teacher groups with medium to below-medium levels of knowledge.

Regarding the role of beliefs, the profile with those teachers who had the highest levels of knowledge and skills (#1) was characterized by only a neutral view on mathematics as a dynamic discipline best learned through inquiry. Overall the four groups did not differ nearly as much with respect to beliefs than to knowledge and skills. This result may be interpreted as an indication that beliefs may be shaped by factors other than teacher education or professional development only. Furthermore, it may be an explanation for the inconclusive state of research [e.g., Simmons et al. 1999; Kunter et al. 2013; Bruckmaier et al. (2016)].

Our results revealed that the interplay of competence facets may be more diverse than it is possible to discover in variable-oriented approaches (e.g., Gold et al. 2013; Stürmer et al. 2013), and thus made it possible to differentiate these results. Due to long testing time for as many as eight competence facets, we had to restrict our sample size. If the results can be replicated in other studies, it would be important to take this teacher heterogeneity into account

in further research and practice. The groups have different training needs and it would make both teacher education and professional development more effective if they could tailor their offerings towards these needs.

As hypothesized, the four teacher groups implemented different levels and types of INQUA where each group displayed differential strengths and weaknesses: teachers with pronounced levels of knowledge and skills (profile #1), as a tendency but not significantly more often teaching at a Gymnasium than at other types of schools, succeeded with respect to student support, cognitive activation and mathematics-related instructional quality, while they implemented only a medium level of classroom management. Teachers with high levels of cognitive skills in M\_PID and CME (profile #3) also succeeded with respect to cognitive activation and mathematics-related quality despite only medium levels of knowledge. However, they struggled with classroom management and student support, which is a surprising result given that CME assesses classroom management expertise. Our assumption is that this result may be affected by the type of schools to which teachers belong. Another explanation could be related to sample bias. CM was on average positively evaluated by raters, which may indicate a positive selection and thus a lack of larger disruption by students. With such a reduction of variance in CM, the chance of finding CME effects might be reduced.

Classroom management was a strength of those teachers belonging to profile #2 characterized by medium levels of all competence facets. However, these teachers struggled with mathematics-related quality, cognitive activation and student support.

The small group of teachers with rather low levels of knowledge and skills except on MSpeed (profile #4), as a tendency teaching more often at non-academic types of schools, struggled with all INQUA facets. We do not have an explanation why this group achieved an almost medium level of MSpeed despite very low MCK. Conceptually, MSpeed requires experience with typical student errors, in order for teachers to be able to discover these fast enough in our time-limited design.

If one looks across the four INQUA facets to identify patterns in their relation to the teacher profiles, it seems as if there may be a difference between classroom management on the one hand and student support, cognitive activation and mathematics-related quality on the other hand. The three latter facets seem to be more closely related to subject-specific knowledge and skills than classroom management. Given the long-standing discussion about the nature of these facets, concerning the extent to which they are generic or subject-specific (Charalambous and Praetorius 2018), this result may move our view on the INQUA facets even further to subject-specificity than before. In the beginning of

research on instructional quality, student support and cognitive activation were almost exclusively regarded as generic ('basic') quality facets that appear across subjects. And even though the corresponding scales and coding schemes were often applied to one specific subject, in particular mathematics, the types of indicators used and their wording was conceptualized from a generic cognitive-psychological point of view. In contrast, more recently not only a fourth subject-specific facet that includes quality criteria of mathematics education, hardly transferable to other subjects, has been added (Schlesinger et al. 2018; Lipowsky and Bleck 2019) but cognitive activation has also been operationalized and phrased in a more subject-specific way (Schlesinger et al. 2018).

The results from the present study indicate that also student support may be subject-specific given that it is significantly related to those teacher profiles characterized by strong subject-specific knowledge and skills. Such an interpretation may at least apply to those studies where student support is operationalized in terms of supporting individual learning needs, for example through differentiation during a lesson, as was done in the present study. It remains an open question whether this also would apply to an operationalization that focuses on fostering classroom climate, which is another way to look at the construct (e.g., Fauth et al. 2014). However, our results will have to be confirmed in other studies first before it is possible to make more definite conclusions, given that we included only a small sample of mathematics teachers and applied an exploratory analysis approach. The results should therefore be taken more as a means to raise awareness of a potential way to think about INQUA, than as a matter of fact.

Before we turn to conclusions, we would like to point out some limitations of our study. Firstly, although  $n = 77$  is a decent sample size given the broad range of instruments to be taken by teachers besides hectic workdays, the size has to be regarded as rather low given the statistical power needed for accomplishing stable and generalizable results. Furthermore, the sample was presumably biased towards a positively selected group of teachers. Overall, it would therefore be important to replicate this study with a larger group of teachers—however difficult it may be to recruit these.

Secondly, the present study was limited to mathematics. It could build on a long research tradition where the different types of instruments were developed and validated with several national and international samples. It would be important to establish similar research traditions with respect to other subjects because it is a relevant question to what extent the profiles identified are generalizable across subjects, including how subject-specific INQUA models are. Policy reforms are rarely implemented for only one subject. But implementing far-reaching reforms of teacher education based on results

from mathematics teachers only may be risky, as for example, studies on the differential results regarding instructional quality facets revealed: while classroom management did not vary significantly across German and English lessons taught by the same teachers, student support did so (Praetorius et al. 2015). Finally, we need to point out that we had to exclude MPCK for statistical reasons to avoid extracting the wrong number of profiles due to residual covariances. The construct is indirectly represented though, due to the conceptual overlap with MCK and M\_PID.

As a first conclusion, we would like to summarize that our results about the relation between the four competence profiles and INQUA may provide a preliminary answer, regarding mathematics teaching, to the questions raised by Blömeke and Kaiser (2017): which level is enough in order for a teacher to be called “competent”? And is it possible to compensate for weakness on one or several competence facets through strengths on others? There is an agreement in the literature that classroom management has to be regarded as an important but not sufficient prerequisite of a teacher’s chance to implement the other INQUA facets. When it comes to student achievement, there is evidence that other facets may be of higher relevance. In this sense the profile of the first class looks most promising. If studies with larger samples replicate such a profile and, furthermore, can link this not only to INQUA but also to student achievement, this would support the need of strong levels of knowledge and skills for successful mathematics teaching. It would at the same time indicate a lesser relevance of a specific belief profile.

The INQUA implemented by teachers belonging to the third profile may indicate, but this second conclusion has to be used with great caution, that strong cognitive skills may be able to compensate at least partly for some weaker spots in teacher knowledge. The final answer to this question will not only depend on replications of our results but also on linking teacher and INQUA characteristics to student outcomes. The state of research with respect to teacher knowledge indicates positive effects either on cognitive outcomes such as mathematics achievement (e.g., Hill et al. 2005, Kunter et al. 2013) or on affective-motivational outcomes such as student motivation. Although these studies are first indications of the relevance of teacher competence and INQUA, results about relations of more recently debated skill facets such as teachers’ perception, interpretation and decision-making to student outcomes are lacking as well as the interplay of the full range of competence facets.

**Acknowledgments** Open Access funding provided by Oslo University & Oslo University Hospital.

**Funding** Funding was provided by Federal Ministry of Education and Research.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Asparouhov, T. & Muthén, B. (2013). Auxiliary variables in mixture modeling: A 3-step approach using Mplus (=Mplus Web Notes, 15). [http://statmodel.com/examples/webnotes/AuxMixture\\_submitted\\_corrected\\_webnote.pdf](http://statmodel.com/examples/webnotes/AuxMixture_submitted_corrected_webnote.pdf). Accessed 22 Jan 2020.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers’ mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180.
- Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education*, 56, 35–46.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (2015a). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223, 3–13.
- Blömeke, S., Hsieh, F.-J., Kaiser, G., & Schmidt, W. (Eds.). (2014). *International perspectives on teacher knowledge, beliefs and opportunities to learn*. Dordrecht: Springer.
- Blömeke, S., & Kaiser, G. (2017). Understanding the development of teachers’ professional competencies as personally, situationally and socially determined. In J. D. Clandinin & J. Husu (Eds.), *International handbook of research on teacher education* (pp. 783–802). London: Sage.
- Blömeke, S., König, J., Suhl, U., Hoth, J., & Döhrmann, M. (2015b). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit von videobasierten Performanztests. *Zeitschrift für Pädagogik*, 61, 310–327.
- Blömeke, S., Suhl, U., & Döhrmann, M. (2012). Zusammenfügen was zusammengehört: Kompetenzprofile am Ende der Lehrerbildung im internationalen Vergleich. [Joining together what belongs together: An international comparison of competence profiles at the end of teacher training]. *Zeitschrift für Pädagogik*, 58, 422–440.
- Bruckmaier, G., Krauss, S., Blum, W., & Leiss, D. (2016). Measuring mathematics teachers’ professional competence by using video clips (COACTIV video). *ZDM Mathematics Education*, 48(1–2), 111–124.
- Carlson, J., & Daehler, K. R. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers’ knowledge for teaching science* (pp. 77–92). Singapore: Springer.

- Charalambous, C. Y., & Praetorius, A.-K. (2018). Studying mathematics instruction through different lenses: Setting the ground for understanding instructional quality more comprehensively. *ZDM Mathematics Education*, 50(3), 355–366.
- Depaepe, F., & König, J. (2018). General pedagogical knowledge, self-efficacy and instructional practice: Disentangling their relationship in pre-service teacher education. *Teaching and Teacher Education*, 69, 177–190.
- Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, 34, 12–25.
- Evertson, C. M., & Weinstein, C. S. (2013). *Handbook of classroom management: Research, practice, and contemporary issues*. New York: Routledge.
- 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.
- Gold, B., Förster, S., & Holodynski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht. *Zeitschrift für Pädagogische Psychologie*, 27, 141–155.
- Goldin, G. A., Hannula, M. S., Heyd-Metzuyanin, E., Jansen, A., Kaasila, R., Lutovac, S., et al. (2016). *Attitudes, beliefs, motivation and identity in mathematics education. An overview of the field and future directions*. Cham: Springer.
- Gräsel, C., Decristan, J., & König, J. (2017). Adaptiver Umgang mit Heterogenität im Unterricht. *Unterrichtswissenschaft*, 45(4), 195–206.
- Hart, S., & Drummond, M. J. (2014). Learning without limits: Constructing a pedagogy free from determinist beliefs about ability. In L. Florian (Ed.), *The Sage handbook of special education* (2nd ed.). London: Sage Publications.
- Heinrichs, H., & Kaiser, G. (2018). Diagnostic competences for dealing with students' errors: Fostering diagnostic competence in error situations. In T. Leuders, K. Philipp, & J. Leuders (Eds.), *Diagnostic competence of mathematics teachers* (pp. 79–94). Cham: Springer.
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte: Information Age.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., et al. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, 26, 430–511.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371–406.
- Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM Mathematics Education*, 48(1–2), 41–53.
- 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.
- Kaiser, G., & König, J. (2019). Competence measurement in (mathematics) teacher education and beyond: Implications for policy. *Higher Education Policy*. <https://doi.org/10.1057/s41307-019-00139-z>.
- Kerstin, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49, 568–589.
- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3–22). Ashland: Hogrefe & Huber Publishers.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster: Waxmann.
- Klieme, E., & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik: Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54, 222–237.
- Knievel, I., Lindmeier, A. M., & Heinze, A. (2015). Beyond knowledge: Measuring primary teachers' subject-specific competences in and for teaching mathematics with items based on video vignettes. *International Journal of Science and Mathematics Education*, 13(2), 309–329.
- Koepfen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modelling and assessment. *Zeitschrift für Psychologie/Journal of Psychology*, 216, 61–73.
- König, J., Blömeke, S., & Kaiser, G. (2015). Early career mathematics teachers' general pedagogical knowledge and skills: Do teacher education, teaching experience, and working conditions make a difference? *International Journal of Science and Mathematics Education*, 13, 331–350.
- 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.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart and Winston.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., et al. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716–725.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105, 805–820.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lee, J., Meadows, M., & Lee, J. O. (2003). *What causes teachers to implement high quality mathematics education more frequently: Focusing on teachers' pedagogical content knowledge*. Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education (ED 472 327).
- Leuders, T., Philipp, K., & Leuders, J. (Eds.). (2018). *Diagnostic competence of mathematics teachers: Unpacking a complex construct in teacher education and teacher practice* (Mathematics Teacher Education; 11). Cham, Switzerland: Springer International Publishing.
- Lipowsky, F., & Bleck, V. (2019). Was wissen wir über guten Unterricht? Ein Update. In U. Steffens & R. Messner (Eds.), *Unterrichtsqualität: Konzepte und Bilanzen gelingenden Lehrens und Lernens* (pp. 219–249). Münster: Waxmann.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean theorem. *Learning and Instruction*, 19(6), 527–537.
- Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. Hoboken: Wiley.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767–778.

- Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*(1), 21–39.
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Muthén, L. K., & Muthén, B. O. (1998–2018). *Mplus user's guide: Statistical analysis with latent variables* (8th edn.). Los Angeles: Muthén & Muthén.
- Nylund, K. L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569.
- Oser, F. (2013). I know how to do it, but I can't do it: Modeling competence profiles for future teachers and trainers. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn, & J. Fege (Eds.), *Modeling and measuring competencies in higher education* (pp. 45–60). Rotterdam: Sense Publishers.
- Pajares, F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*, 307–332.
- Pankow, L., Kaiser, G., König, J., & Blömeke, S. (2018). Perception of student errors under time limitations: Are teachers faster than mathematicians or students? *ZDM Mathematics Education, 50*, 631–642.
- Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257–315). Charlotte: Information Age Publishing.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM Mathematics Education, 50*(3), 407–426.
- Praetorius, A.-K., Vieluf, S., Saß, S., Bernholt, A., & Klieme, E. (2015). The same in German as in English? Investigating the subject-specificity of teaching quality. *Zeitschrift für Erziehungswissenschaft, 19*(1), 191–210.
- Ramaswamy, V., DeSarbo, W., Reibstein, D., & Robinson, W. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*, 103–124.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), *Handbook of research on teacher education* (pp. 102–119). New York: Macmillan.
- Rowland, T., & Ruthven, K. (Eds.). (2011). *Mathematical knowledge in teaching*. Dordrecht: Springer.
- Santagata, R., & Guarino, J. (2011). Using video to teach future teachers to learn from teaching. *ZDM - The International Journal on Mathematics Education, 43*(1), 133–145.
- Sass, D., Schmitt, T., & Marsh, H. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(2), 167–180.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference Chi square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507–514.
- Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM Mathematics Education, 48*(1), 29–40.
- Schlesinger, L., Jentsch, A., Kaiser, G., König, J., & Blömeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM Mathematics Education, 50*, 475–490.
- Schoenfeld, A. H. (2010). *How we think*. New York: Routledge.
- Schön, D. A. (1983). *The reflective practitioner. How professionals think in action*. New York: Basic Books.
- Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.
- Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York: Routledge.
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.
- Simmons, P. E., Emory, A., Carter, T., Coker, T., Finnegan, B., Crockett, D., et al. (1999). Beginning teachers: Beliefs and classroom actions. *Journal of Research in Science Teaching, 36*, 930–954.
- 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*(2), 107–125.
- Stürmer, K., Könings, K. D., & Seidel, T. (2013). Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *British Journal of Educational Psychology, 83*, 467–483.
- Stürmer, K., & Seidel, T. (2015). Assessing professional vision in teacher candidates. Approaches to validating the observer extended research Tool. *Zeitschrift für Psychologie, 223*(1), 54–63.
- Südkamp, A., & Praetorius, A.-K. (2017). *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen*. Münster: Waxmann.
- Tatto, M. T., Schulle, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework*. East Lansing: Michigan State University.
- Tatto, M. T., Schulle, J., Senk, S., Ingvarson, L., Rowley, G., Peck, R., et al. (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)*. Amsterdam: IEA.
- Voss, T., Kleickmann, T., Kunter, M., & Hachfeld, A. (2013). Mathematics teachers' beliefs. In M. Kunter, et al. (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers (Mathematics Teacher Education, 8)* (pp. 249–271). New York: Springer.
- Voss, T., Kunter, M., Seiz, J., Hoehne, V., & Baumert, J. (2014). Die Bedeutung des pädagogisch-psychologischen Wissens von angehenden Lehrkräften für die Unterrichtsqualität. *Zeitschrift für Pädagogik, 60*(2), 184–201.
- Weinert, F. E. (2001). Concepts of competence: A conceptual clarification. In D. S. Rychen & L. H. Salgnik (Eds.), *Defining and selecting key competencies* (pp. 45–66). Göttingen: Hogrefe and Huber.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-aspect test software*. Camberwell: Australian Council for Educational Research.
- Yang, X., Kaiser, G., König, J., & Blömeke, S. (2018). Professional noticing of mathematics teachers: A comparative study between Germany and China. *International Journal of Science and Mathematics Education*. <https://doi.org/10.1007/s10763-018-9907-x>. (Advance online publication June 19).
- Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*(1), 165–200.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.