Opening up the black box: Teacher competence, instructional quality, and students' learning progress

Sigrid Blömeke [a,*], Armin Jentsch [b], Natalie Ross [c], Gabriele Kaiser [c], Johannes König [d]

[a] *Centre for Educational Measurement (CEMO), University of Oslo, Norway*
[b] *Faculty of Philosophy, University of Greifswald, Germany*
[c] *Faculty of Educational Sciences, University of Hamburg, Germany*
[d] *Faculty of Human Sciences, University of Cologne, Germany*

ARTICLE INFO

ABSTRACT

Existing research indicates inconsistent or at best weak predictive effects of teacher knowledge on student achievement. Data from Germany were used to examine the relation between teachers' content and pedagogical content knowledge, their perception, interpretation, and decision-making skills, the instructional quality implemented in class, and students' learning progression in mathematics. Rather than direct effects of teacher knowledge on students, we hypothesized an effect chain with multiple mediation processes while controlling for school type and student background. Multi-level modeling with 3496 students from 154 classrooms revealed a mediating role of teachers' skills and their instructional quality for the relation between teacher knowledge and students' learning progress. Effect sizes were medium to strong, and the model explained a large amount of variance. No direct effects of teachers' knowledge on student progress were found. We discuss our findings with respect to the teacher-competence-as-a-continuum model and with respect to future research.

1. Introduction

The role of teacher competence in students' learning progression has become a key topic in educational research. This is particularly the case in the field of mathematics education (Kaiser & König, 2020). The quality of corresponding studies has increased substantially over the past years, yielding results that are more reliable and valid. Studies have progressed beyond self-reports of teachers' competence by including standardized tests of different competence facets and relating these to student achievement (e.g., Baumert et al., 2010; Hill et al., 2005; Kersting et al., 2012). Other studies have examined the role that teacher competence plays in the quality of instruction delivered in the classroom (e.g., Charalambous & Praetorius, 2018; Jentsch et al., 2021) and progressed beyond cross-sectional designs by implementing a longitudinal component at the student level to assess the relation between learning progression and selected facets of teacher competence (e.g., Kunter et al., 2013).

However, this research has not yet managed to establish a robust link between teacher competence and student progress. Effect sizes varied between weakly positive and weakly negative estimates (Blömeke &

Olsen, 2019; Seidel & Shavelson, 2007) with most studies revealing no effects at all, particularly with respect to the relation of teachers' content knowledge as one competence facet to student outcomes. However, a major limitation of this research was that only few studies have examined the full effect chain, including a broad range of teacher competence facets, instructional quality, and students' learning progress within one study. Most studies were either limited in scope with respect to the range of competence facets assessed, treated the relation between competence and student achievement as a "black box" by omitting instructional quality as Baumert et al. (2010) described it (see also Hill et al., 2005), or implemented a cross-sectional design (e.g., Blömeke & Olsen, 2019).

Against this background, the purpose of the present paper is to examine the relation between teacher competence, instructional quality, and students' learning progression more comprehensively. This applies, first, to the range of cognitive competence facets included and, second, to the mediating processes which may transform teacher competence into student progress.

---

2. Theoretical framework

2.1. Teacher competence

During the last two decades, scholars have established elaborate models of teacher competence that conceptualize a broad range of cognitive and affective-motivational characteristics teachers need to successfully perform their work (e.g., Baumert & Kunter, 2006; for an overview see Blömeke & Kaiser, 2017). To promote the development of domain-specific student cognitions, these models typically stress the relevance of domain-specific cognitive competence facets, more precisely of teachers' domain-specific knowledge on the one hand and their domain-specific cognitive skills to perceive classroom situations, to interpret these and to make decisions on the other hand (Kaiser et al., 2017). These two types of cognitive competence facets are therefore the focus of the present study. Affective-motivational teacher characteristics are also regarded as important for domain-specific student outcomes but are not at the forefront of this study.

Based on Shulman's seminal work (1986), teachers' domain-specific *knowledge* can be divided into content knowledge and pedagogical content knowledge. In the case of mathematics teachers, the former denotes mathematical content knowledge (MCK) and the latter denotes mathematics pedagogical content knowledge (MPCK). MCK includes knowledge about content domains, such as number, algebra, geometry, and data, which should provide teachers with the necessary background knowledge for teaching. MPCK covers curricular knowledge and planning for mathematics teaching and enacting, which should provide teachers with the necessary knowledge for in-class lesson activities (e.g., Tatto et al., 2008; Kunter et al., 2013). Empirical studies have demonstrated that these facets constitute two distinct knowledge dimensions (e.g., Krauss et al., 2008; Baumert et al., 2010; Blömeke et al., 2016).

Teachers' domain-specific cognitive *skills* are organized in line with classroom situations with particular emphasis on situations that are decisive for students' domain-specific learning progression, more precisely with emphasis on those characteristics important for high quality in these situations, such as their instructional design, the potential for students' cognitive activation, individual learning support, and classroom management (Blömeke et al., 2015). A crucial distinction between teachers' knowledge and skills is their proximity to observable behavior in class. Whereas knowledge comprises generalized cognition not necessarily related to one specific classroom situation, cognitive skills are typically organized in a context-related way (Putnam & Borko, 2000). It is important to consider that these skills are conceptualized as *cognitive* teacher characteristics and are not equal to observed behavior, which is discussed in the next section called "instructional quality". Cognitively, teachers need to perceive a specific classroom situation as relevant and interpret its different aspects to be able to determine how to act, for example, by anticipating potential student responses or developing alternative instructional strategies (cf. Jacobs et al., 2010). Studies indicate that teachers' cognitive skills can be described as two-dimensional, with one generic and one domain-specific facet (Blömeke et al., 2016). The latter facet is focused in the present study; its three components—perception, interpretation, and decision-making (PID)—have been identified being as interrelated due to their process-oriented character, to the extent that they are difficult to disentangle (Santagata & Yeh, 2016; Stahnke et al., 2016).

2.2. Instructional quality

Instructional quality reflects teachers' observable classroom behavior and can be defined by three dimensions conceptualized to some extent independently of a specific teaching domain—namely classroom management, cognitive activation, and student support (Klieme et al., 2009; Praetorius et al., 2018)—and domain-specific

quality characteristics which are, in case of our study, related to mathematics education (Schlesinger et al., 2018). In a literature review, Charalambous and Praetorius (2018) compiled evidence that predicting students' learning outcomes in mathematics based on domain-specific and generic characteristics of instructional quality may be more successful than when predicted based solely on generic characteristics. Empirical studies applying different observation instruments support this comprehensive approach to modeling INQUA (Kane & Staiger, 2012). Inferring from the way how these are conceptualized in the literature and depending on the breadth of characteristics included, this domain-specific quality can be modelled as consisting of one or several dimensions (Kyriakides et al., 2013; Seidel & Shavelson, 2007).

Models of instructional quality typically conceptualize classroom management as the efficient use of allocated classroom time, the prevention of disorder in the classroom, and the clarity of organizational rules (Praetorius et al., 2018). Diagnosis of students' learning, provision of opportunities for individualization and differentiation, and the creation of a good teaching climate are subsumed under the notion of student support (Fauth et al., 2014). Cognitive activation refers to whether students are challenged by higher-order thinking through teachers' instructional strategies and the learning tasks selected (Klieme et al., 2009; Lipowsky et al., 2009). In our study, educational quality in mathematics is characterized by an appropriate presentation of the content and content-specific interaction with students, including the provision of domain-specific feedback (Schlesinger & Jentsch, 2016; Jentsch, Schlesinger, Heinrichs, Kaiser, König, & Blömeke, 2021).

2.3. Students' learning progression

The framework underlying national and state-wide tests of student achievement in school mathematics in Germany encompasses the so-called "big ideas" that form a three-dimensional model (Blum et al., 2006). The content assessed comprises numbers, measurement, space and form, functional relations, and data and chance. The processes assessed comprise argumentation, problem solving, modeling, representations, using symbolic, formal, and technical elements, and communication. These dimensions are organized along three levels of achievement: reproduction, making connections, and generalization and reflection.

2.4. Relations between teacher competence, instructional quality, and learning progression

Given the weak correlations identified hitherto between teacher knowledge and student achievement, Blömeke et al. (2015) proposed a more elaborate model hypothesizing potential mediating processes by drawing on insights from cognitive psychology. Rather than treating the relation between teacher knowledge and student achievement as a "black box", they considered it crucial to evaluate how distal or proximal teacher and teaching characteristics were to a desired outcome.

The model by Blömeke et al. (2015) conceptualized teachers' knowledge facets as traits that are relatively stable across different situations, representing the potential that a teacher brings to the classroom. This part of the model is similar to the teacher knowledge models developed, for example, by Ball et al. (2008) or Baumert and Kunter (2006).

The model by Blömeke et al. (2015) also considers Schön's (1983) concepts of reflection in and on action by including cognitive processes immediately prior to, during, and following typical classroom situations (Star & Strickland, 2008). Teachers' skills to perceive, interpret and make decisions, that is their PID skills, were thus conceptualized as a context-related and situated facet of their cognitions, equal neither to more generalizable knowledge nor to observable teaching behavior in specific classroom situations. This model thereby merged the previously

discrete understandings of teacher competence from a dispositional or a situated perspective (Rowland & Ruthven, 2011). PID skills were thus conceptualized as mediators between teacher knowledge and teaching behavior in the classroom and functionally related to both. Likewise, Meschede et al. (2017, p. 159) argued that teachers' skills were an "essential mediator" between teachers' knowledge and their teaching.

Observable teaching behavior was conceptualized as being most closely related to student outcomes. Summing up, Blömeke et al. (2015) suggested considering an effect *chain* that acknowledges teachers' PID skills as an important mediator in the transformation of teacher knowledge into teaching behavior and this teaching behavior as another mediator into students' learning progression.

Whereas this theory argues well for a potentially causal chain—knowledge–skills–behavior and outcomes—some parts of the model are less elaborate. This applies first to the relationship between different knowledge dimensions and teachers' PID skills, and second to direct versus indirect effects. Blömeke et al. (2015) leave it open how exactly MCK and MPCK are related to PID skills. MCK is less clearly related to mathematics teaching than MPCK as MCK includes general mathematical knowledge that shall serve as foundational background knowledge whereas MPCK is related to teaching tasks (for examples see sections 3.1 and 3.2 in the electronic supplement). The legacy of research on cognitive abilities points in such cases to placements at different hierarchical levels. More general abilities are assumed to underly more specific ones (Jensen, 1998). A reasoning in line with such a model would mean that MPCK would be hypothesized to have direct effects on teachers' PID skills while MCK would be hypothesized to be predictive for MPCK (cf. similar discussions by Ball et al., 2008; Baumert et al., 2010).

The description by Blömeke et al. (2015) gives in addition rise to assuming indirect effects of teacher knowledge on instructional quality and students' learning progression only. However, instead of hypothesizing that effects of MCK and MPCK on teaching behavior or student outcomes are *fully* mediated by teachers' PID skills, which is a very strict requirement, direct effects would–at least to some extent–be plausible as well. With respect to the relation between teacher knowledge and instructional quality, such direct effects may reflect a richer domain-specific terminology, for example.

3. State of research

3.1. Relation of MCK and MPCK to teachers' PID skills

As noted, studies intending to establish direct relations between mathematics teachers' content knowledge and students' learning progression without considering potentially mediating characteristics of the hypothesized effect chain knowledge–skills–teaching behavior–outcomes typically found no or weak predictive effects (e.g., Hanushek et al., 2005; Hill et al., 2007). Direct effects of MPCK on student achievement have rarely been examined: we identified one study where a significant medium (according to Cohen, 1988; see this reference also in the following for categorization of effect sizes) effect was found for German lower-secondary mathematics teachers (Baumert et al., 2010).

More studies have examined small parts of the hypothesized effect chain, identifying systematic relations more successfully. The examination of the relation between MCK and MPCK has predominantly revealed strong correlations (e.g., Tatto et al., 2012). Regarding their relation to teachers' PID skills to perceive, interpret and make decision and potential causality, Dunekacke et al. (2015) demonstrated in the context of German early childhood education and care that MCK was a necessary precondition for MPCK. Baumert et al.'s (2010) findings with respect to lower-secondary mathematics teachers can be interpreted similarly.

In both projects, teachers' PID skills were also assessed (though differently operationalized). The data revealed that both teachers' MCK

and MPCK were significantly positively related to these skills with a stronger effect size for MPCK than for MCK (Blömeke et al., 2016; Bruckmaier et al., 2016). A significant relation between both MCK and MPCK and skills—focused on mathematical representations—was also identified in Dreher and Kuntze's (2015) study among German lower-secondary teachers. However, the effects were inconsistent for different groups of teachers with MCK showing significant effects for student teachers but MPCK for practicing teachers. US studies of mathematics teachers revealed significant relations between different types of their content knowledge and their skills but did not include pedagogical content knowledge (Kersting et al., 2012; Hill & Chin, 2018). Overall, existing research indicates a positive relation between MCK and MPCK and teachers' PID skills, but the exact nature of its interplay remains unclear.

3.2. Relation between teacher competence and instructional quality

The relation between the different facets of mathematics teachers' competence and the instructional quality implemented in class has also been examined in a range of studies, albeit mostly limited to one or two competence facets and some facets of instructional quality. Baumert et al. (2010) found a significant relation of medium effect size between MCK and a domain-specific dimension of instructional quality—curricular alignment—but not to the three generic dimensions. MPCK was related to cognitive activation, also with a medium effect size, but not to curricular alignment, student support, or classroom management (see also Kunter et al., 2013).

Hill et al. (2007) provided evidence with respect to primary education and the domain-specific characteristics of instructional quality as follows: Teachers with lower content knowledge made more mathematical errors in their instruction, while teachers with higher content knowledge used richer representations, explanations, and justifications. Effect sizes were medium to large. Kelcey et al. (2019) replicated these findings. Bruckmeier et al. (2016) also demonstrated that teachers' skills were related to instructional quality.

Kersting et al. (2012) conducted a study that included both a paper-pencil test of mathematics teachers' knowledge and an assessment of teachers' cognitive skills where they had to answer questions related to video-cued classroom situations, both focused on fractions. They found no direct effects of the results of the paper-pencil test on instructional quality but found a relation of medium effect size between the two instruments and an effect of large size of the video-based measure on instructional quality when both measures were included. This may indicate that only teachers' skills are related to instructional quality, while MCK may have an indirect effect on it via teachers' skills. Overall, it seems plausible to infer an effect of teachers' knowledge and skills on instructional quality from such findings, but the precise interplay remains unclear.

3.3. Relation between instructional quality and students' learning progression

Finally, studies have examined the relation between instructional quality and students' learning progression. Data from the German-Swiss "Pythagoras" study (Lipowsky et al., 2009) revealed positive effects for both classroom management and cognitive activation of small size. Hill et al. (2007), Kersting et al. (2012), and Kelcey et al. (2019) showed that higher quality instruction positively affected student learning gains and mediated the role of teachers' content knowledge. Effect sizes varied in these studies. In the study by Kunter et al. (2013) high levels of cognitive activation and efficient classroom management were also found to mediate the effect of MPCK and to predict greater achievement gains. Overall, the state of research is consistent in treating instructional quality as a mediator of knowledge effects on student progression. In particular, cognitive activation and effective classroom management are those basic dimensions of instructional quality that contribute to

learning gains among students.

3.4. Potentially confounding variables

To estimate teacher effects on student outcomes correctly, it is important to control for potentially confounding variables that would otherwise create bias in the results due to non-random student–teacher assignments (Koedel et al., 2015). Reviews of studies in the context of educational effectiveness research and value-added modeling point to prior achievement, student background, and school context as core variables that were significantly related to outcomes—mostly with large effect sizes. They should therefore be included in studies to ensure fair comparisons (Levy et al., 2019). Furthermore, most studies mentioned above revealed confounding effects of student background and—in Germany, particularly—school type (e.g., Baumert et al., 2010), with large effect sizes.

4. Research questions and hypotheses

The state of research allows for deriving directional hypotheses, in a few cases also including conjectures on the degree of the effect size based on the reliability and the expected proximity of the constructs assessed compared to other studies. However, in most cases it is impossible to infer the potential effect size from the literature due to greatly varying study designs and results. Based on the research described above, the relations between teacher competence, instructional quality, and students' learning progression are hypothesized as follows (see Fig. 1):

1) Teachers' MCK has a positive predictive effect on MPCK (H1a). MPCK has a positive predictive effect on teachers' PID skills related to mathematics instruction (H1b). These skills are expected to at least partly mediate the effect of MCK (H1c). MCK may or may not have an additional direct effect on teachers' skills.
2) It is then hypothesized that mathematics teachers' PID skills predict students' learning progression in mathematics (H2a) and that the effects of MCK and MPCK are at least partly mediated by these skills (H2b). MCK and MPCK may or may not additionally directly affect student progress.
3) Having clarified the relation of MCK and MPCK to teachers' PID skills and students' progress, teacher competence is hypothesized to be positively related to instructional quality (H3a), which in turn significantly positively affects students' learning progression in mathematics (H3b). Furthermore, instructional quality is hypothesized to play a mediating role regarding the effects of teacher competence and student progress (H3c). Teachers' PID skills, MCK or MPCK may or may not have direct additional effects on student progress.

5. Methodology

In the following, the context of the study, the sample, psychometric properties of all instruments including item examples, and the data analysis including results of the different robustness checks are described briefly. For details see the electronic supplement.

5.1. Context of the present study: TEDS-Instruct and TEDS-Validate

The data stem from two studies— TEDS-Instruct and TEDS-Validate —conducted in Germany within a research program departing from the international "Teacher Education and Development Study in Mathematics" (TEDS-M) in which MCK and MPCK tests were developed (Tatto et al., 2008). In a German follow-up study (TEDS-FU; Kaiser et al., 2015), Blömeke et al.'s (2015) broader teacher competence framework was applied; and tests of mathematics teachers' PID skills were developed. TEDS-Instruct, carried out in the federal state of Hamburg, and TEDS-Validate, carried out in the federal state of Thuringia, added tests of students' learning progression (Kaiser & König, 2020). Both states have two types of middle school: an academic track called *Gymnasium* and a non-academic track called *Stadtteilschule* (Hamburg), *Regelschule*, or *Gesamtschule* (Thuringia).

5.2. Sample

Our sample comprised 3496 students from 154 classrooms. Of the students, 51.2% were female, 15.8% did not speak German as their first language, and 4.8% had special needs. Average class size was 22.7 students. In the federal state of Hamburg, the classes sampled were tested at the beginning of grade 7 in 2012, 2013, and 2014 and followed up about 1.5 years later at the end of grade 8. In the federal state of Thuringia, the classes sampled were tested at the end of grade 6 in 2012, 2013, 2014, 2015, and 2016 and followed up about two years later at the end of grade 8. The 154 classes were taught by 89 teachers. About half of the teachers were female, and 63.2% were teaching at a Gymnasium. On average, they were 40 years old and had been teaching for 13 years.

Teachers' MCK, MPCK, and PID skills were assessed simultaneously during one computer-based session that could be paused once. Observations of teachers' instructional quality were available for 49 teachers because this element had not been included in all studies. Since the vast majority of these teachers (*n* = 37, teaching 879 students) were from Hamburg, we restricted the models including instructional quality to Hamburg to avoid an unbalanced sample. Teachers volunteered to participate in the study. Therefore, we must assume self-selection bias.

5.3. Measures

5.3.1. Teacher competence
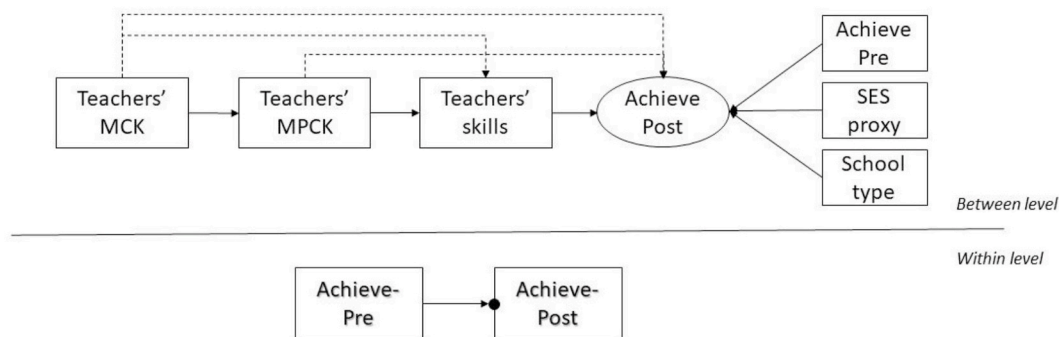Mathematics teachers' competence was assessed with digitalized

Fig. 1. Mediation model of the relation between teachers' domain-specific knowledge, their PID skills and students' learning progression in mathematics, controlling for school type and student background (dotted lines: competing hypotheses of direct versus indirect effects of MCK and MPCK).

tests that covered MCK, MPCK, and PID skills (Kaiser et al., 2015). Scaled scores were created by applying item-response theory as implemented in the software package Conquest (Wu et al., 1997). Items omitted or not reached were considered as incorrect responses.

MCK was assessed internet-based with an abbreviated version of the original paper-and-pencil TEDS-M test (Tatto et al., 2012), validated in several studies (see Blömeke et al., 2016 for an overview). The 27 items included numbers, algebra, data, and geometry as core areas of school mathematics. Reliability both in terms of Cronbach's α and Warm's weighted likelihood estimate (WLE) was 0.81. MPCK was assessed with the original TEDS-M test (Tatto et al., 2012). The 28 items covered curricular and planning knowledge as well as knowledge about how to teach mathematics. Reliability was α = 0.79 or WLE = 0.78.

Teachers' PID skills were assessed video-based with 32 items requiring them to perceive, interpret, and make decisions with respect to typical classroom situations presented in three scripted video clips (Kaiser et al., 2015). The clips served as cues and lasted between 2.5 and 4 min. The scale's reliability was α = 0.72 or WLE = 0.73.

5.3.2. Instructional quality

Instructional quality was assessed using a standardized observation protocol (Schlesinger et al., 2018). The 21 items covered classroom management (4 items), student support (5), cognitive activation (6), and mathematics educational structuring (6) as indicators of one latent variable to reduce complexity, given that we examined the entire effect chain from teacher knowledge to student achievement, and to avoid multicollinearity. Confirmatory factor analysis revealed a good fit of the data to such a model (CFI = 0.96, RMSEA = 0.02, $SRMR_{BETWEEN}$ = 0.04, X^2 = 2.6, df = 2, p = .27).

5.3.3. Student achievement in mathematics

Student achievement at the first measurement point was assessed by state-wide tests developed in the federal states of Hamburg and Thuringia based on the German national standards. At the second measurement point, student achievement was assessed by a national test based on these nation-wide standards. The data in terms of person parameters (WLE) resulting from item-response-theory scaling were provided by the respective units responsible for testing in the two federal states. Since the data came from different cohorts, we refrain from comparing absolute values across groups. Since the period between pre- and post-tests also varied (1.5–2 years), we also refrain from reporting students' absolute learning gains.

5.3.4. Control variables

As an indicator of students' average educational background, we used their language background. Since this information was not available at the individual level, we included the proportion of students with German as their first language at class level, being aware that the meaning of this variable may differ as a function of aggregation (see the Limitations section). Given the differences in learning opportunities between the *Gymnasium* and the non-academic track in Germany, the type of school attended was included as a school context variable.

5.4. Data analysis

We applied a series of two-level random-intercept mediation models with students on the first and classes on the second level to test our hypotheses. To account for the difference in the number of classes taught by one teacher, we implemented a teacher weight that was inversely proportional to the number of classes taught. Variables were group-mean centered at the within and remained uncentered at the between level. Missing data were handled using the full information maximum likelihood procedure. To examine the robustness of our model, we applied several alternative approaches. The differences were negligible. All analyses were conducted using the statistical software package Mplus version 8 (Muthén & Muthén, 1998-2017–2017).

Direct and indirect effects were estimated. Based on Cohen (1988), we interpreted coefficients around 0.10 as weak, around 0.30 as moderate, and around 0.50 or larger as strong direct effects. In its squared version, a small indirect effect size is around 0.01, medium 0.09, and large 0.25.

6. Results

6.1. Descriptive results and direct effects of teacher knowledge

The intra-class correlation of students' mathematical achievement was 0.54, indicating large differences between classrooms. The predictive effect of achievement at the first time point for results at the second time point was high (β = 0.88) on the between-level, and results from the pre-test explained more than three-quarters of the variance in the post-test results (see Model 1 in Table 1).

School type was strongly related to students' learning progression in mathematics (β = 0.46), while the effect of prior achievement decreased substantially (β = 0.52; see Model 2 in Table 1). Student background—strongly correlated with school type (see Table 2 in the electronic supplement)—had no separate predictive effect on learning progress (β = -.00). As expected, no significant direct effect of teachers' MCK (β = 0.03) or MPCK was observed on students' learning progress (β = -.05; see Model 3 in Table 1, "black-box" model).

6.2. Relation of teacher competence to students' learning progression (H1, H2)

We next tested the relation between teacher competence and students' learning progression in mathematics. Teachers' PID skills were hypothesized to mediate at least partly the effects of teacher knowledge on student progress (H2), while MCK was hypothesized to be a predictor of MPCK, which in turn should predict teachers' PID skills (H1, see Fig. 1). Since student background did not contribute beyond school type, we omitted this control variable in favor of more parsimonious models given small sample size.

Model 4 (see Table 2) reflected MCK's effect on teachers' PID skills in perceiving and interpreting classroom events and deciding how to proceed mediated by their MPCK (H1c). This model was supported by the data. MCK had a strong predictive effect on MPCK (H1a, β = 0.69), which in turn had a strong effect on PID skills (H1b, β = 0.50).

Teachers' PID skills had a significant but weak predictive effect on students' learning progression in mathematics (β = 0.12). They fully mediated the effects of MPCK (β = 0.06) on student progress. This means that there were no direct effects of MPCK on student progress. The same applied to MCK. With respect to this knowledge dimension, MPCK and teachers' PID skills combined fully mediated its effect on learning progression (β = 0.04). Mediation through MPCK (-.07) or skills (0.01) only was insignificant. Since the direct effects of teacher knowledge on student progress were insignificant (MCK = 0.05, MPCK = -.10), the total effects of MCK (0.03 in model 4) and MPCK (-.05) were small and insignificant. Similar to the basic models, school type was an important variable that needed to be controlled for (0.45).

6.3. Testing the full effect chain with instructional quality as another mediator (H3)

Finally, we included instructional quality as another mediator of teacher competence's effect on students' learning progression in mathematics. This analysis was, as already mentioned, restricted to the sample from the federal state of Hamburg. Given the small number of classrooms, we proceeded with a parsimonious model in which (the previously insignificant) direct effects of MCK on teachers' PID skills or student progress or of MPCK on student progress were no longer modelled. Given the inconclusiveness regarding the role of teachers' skills, we tested competing models, one with indirect effects only

Table 1

Basic two-level random intercept models (standardized coefficients, standard errors).

Fixed effects	M1			M2			M3		
	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Within level (N)	2565			2500			2489		
Student achievement T2 on									
Prior achievement	.65	.01	<.001	.65	.01	<.001	.65	.01	<.001
R^2 *Student achievement T2*	.42	.02	<.001	.42	.02	<.001	.42	.02	<.001
Between level (k)	139			136			135		
Student achievement T2 on									
Prior achievement	.88	.02	<.001	.52	.07	<.001	.51	.07	<.001
School type				.46	.07	<.001	.46	.07	<.001
Language background				-.00	.03	.97	.01	.03	.88
MCK							.03	.05	.54
MPCK							-.05	.06	.42
R^2 *Student achievement T2*	.78	.04	<.001	.87	.03	<.001	.87	.03	<.001

Note. MCK = mathematics content knowledge, MPCK = mathematics pedagogical content knowledge, T = time point.

Table 2

Relation of teacher competence to students' learning progression (two-level random intercept model; standardized coefficients, standard errors).

Fixed effects	M4		
	Estimate	SE	p
Within level (N)	2917		
Student achievement T2 on			
Student achievement T1	.65	.01	<.001
R^2	.42	.02	<.001
Between level (k)	139		
Student achievement T2 on			
Student achievement T1	.52	.07	<.001
School type	.45	.07	<.001
Teachers' MCK	.05	.06	.40
Teachers' MPCK	-.10	.06	.11
Teachers' PID skills	.12	.04	.005
Teachers' PID skills on			
Teachers' MCK	.05	.12	.67
Teachers' MPCK	.50	.11	<.001
Teachers' MPCK on			
Teachers' MCK	.69	.05	<.001
MCK with MPCK	–		
MCK$_{ind}$ via PID skills	.01	.02	.68
MCK$_{ind}$ via MPCK	-.07	.04	.11
MCK$_{ind}$ via MPCK and PID skills	.04	.02	.008
MCK$_{tot}$	.03	.04	.51
MPCK$_{ind}$ via PID skills	.06	.02	.007
MPCK$_{tot}$	-.05	.06	.44
R^2 *Teachers' MPCK*	.47	.06	<.001
R^2 *Teachers' PID skills*	.29	.08	<.001
R^2 *Student achievement T2*	.87	.02	<.001

Note. MCK = mathematics content knowledge, MPCK = mathematics pedagogical content knowledge, PID = perception, interpretation, and decision-making, T = time point.

regarding this competence facet (see Model 5 in Table 3) and one with an additional direct effect on students' learning progress (Model 6 in Table 3). Since school type was not significantly correlated with instructional quality (see electronic supplement, Table 2), we estimated both models twice, once without (Models 5a and 6a in Table 3) and once including the control variable (Models 5b and 6b in Table 3; for the fully saturated model see Model 7 in supplement Table 5 in the electronic supplement; differences in the results are negligible.)

Similar to the models reported above, teachers' MCK significantly and with a large effect size predicted their MPCK in all models (0.81). MPCK in turn significantly and with a large effect size predicted teachers' PID skills (0.78). Furthermore, as hypothesized (H3a), teachers' PID skills significantly and with a large effect size predicted instructional quality in all models (0.53–0.58). The remaining results

varied depending on the model.

Model 5a (see Table 3) reflected a model with indirect effects of teacher competence on student progress without controlling for school type. As hypothesized (H3b), instructional quality significantly predicted students' learning progression in this model (0.18). The indirect effects of all three facets of teacher competence were also significant and of medium effect size (PID skills: 0.10; MPCK: 0.08; MCK: 0.07).

The picture changed when a direct effect of teachers' PID skills on student progress was included in the model (see Model 6a in Table 3). This effect was significant (0.17), while the effect of instructional quality disappeared. In line with these results, the indirect effects of teacher knowledge on students' learning progress were significant via teachers' PID skills but not via instructional quality (MCK: 0.10; MPCK: 0.13). The total effects of teachers' competence facets were also significant with somewhat larger effect sizes than in the purely indirect model (MCK: 0.13; MPCK: 0.15; PID skills: 0.20). If indirect paths from a facet of teacher competence to students' learning progress included instructional quality, none of the effects was significant (MCK: 0.02; MPCK: 0.02; PID skills: 0.03).

When we controlled for school type, the relation between instructional quality and student progress in mathematics disappeared in the model including indirect effects of teacher competence only (Model 5b in Table 3) and in the model including a direct effect of teachers' PID skills on students' learning progression (Model 6b in Table 3). In the latter case, the indirect effects of teachers' knowledge via their skills and the total effects of the three facets of teacher competence were no longer significant either.

7. Discussion

The purpose of the present study was to model the relation between teacher knowledge and students' learning progression as an effect chain with mediating processes. Based on Blömeke et al.'s (2015) conceptual framework, these mediating processes should be more proximal to how students learn than teachers' knowledge. Thereby, we intended to refine the state of research that had previously identified only weak or no systematic relations between teacher knowledge and student progress. To achieve this, we utilized data from two German studies that provided a decent sample size for such complex modeling, namely TEDS-Instruct and TEDS-Validate (Kaiser & König, 2020).

In line with existing research, we found no direct effect of teachers' MCK or MPCK on students' learning progression in mathematics, indicating that a "black-box" model (Baumert et al., 2010, pp. 160–161) omitting potentially mediating processes was unable to explain student progress. Although many earlier studies had sought to establish such a distal influence, this result is unsurprising from a cognitive-psychological perspective (Anderson, 1983). The more

Table 3

Relation of teacher competence to instructional quality and student progression (two-level random intercept model; standardized coefficients, standard errors).

Fixed effects	M5a			M5b			M6a			M6b		
	Est.	SE	p	Est.	SE	p	Est.	SE	p	Est.	SE	p
Within level (N)	879			879			879			879		
Student achievement T2 on												
Student achievement T1	.64	.02	<.001	.64	.02	<.001	.64	.02	<.001	.64	.02	<.001
R^2	.41	.02	<.001	.41	.02	<.001	.41	.02	<.001	.41	.02	<.001
Between level (k)	37			37			37			37		
Student achievement T2 on												
Student achievement T1	.92	.03	<.001	.53	.13	<.001	.90	.03	<.001	.57	.13	<.001
School type	–			.45	.14	.001	–			.40	.14	.003
Instructional quality	.18	.09	.04	.07	.09	.44	.06	.08	.41	.03	.09	.77
Teachers' PID skills	–			–			.17	.08	.04	.08	.06	.19
Instructional quality on												
Teachers' PID skills	.58	.10	<.001	.54	.10	<.001	.53	.11	<.001	.53	.10	<.001
Teachers' PID skills on												
Teachers' MPCK	.78	.06	<.001	.78	.06	<.001	.78	.06	<.001	.78	.06	<.001
Teachers' MPCK on												
Teachers' MCK	.81	.05	<.001	.81	.05	<.001	.81	.05	<.001	.81	.05	<.001
Indirect/total effects: Student achievement T2 on												
MCK$_{tot}$	.07	.04	.10	.02	.03	.45	.13	.05	.01	.06	.04	.14
MPCK$_{tot}$	.08	.05	.09	.03	.04	.45	.15	.06	.007	.07	.05	.14
PID skills$_{tot}$	.10	.06	.08	.04	.05	.45	.20	.07	.003	.09	.06	.13
MCK$_{ind}$ via MPCK, PID skills							.10	.06	.06	.05	.04	.20
MCK$_{ind}$ via MPCK, PID skills, instructional quality							.02	.03	.40	.01	.03	.77
MPCK$_{ind}$ via PID skills							.13	.07	.05	.06	.05	.20
MPCK$_{ind}$ via PID skills, instructional quality							.03	.03	.40	.01	.04	.77
PID skills$_{ind}$ via instructional quality							.03^s	.04	.40	.01	.05	.77
R^2 Teachers' MPCK	.66	.08	<.001	.66	.08	<.001	.66	.08	<.001	.66	.08	<.001
R^2 Teachers' PID skills	.60	.10	<.001	.60	.10	<.001	.60	.10	<.001	.60	.10	<.001
R^2 Instructional quality	.33	.12	.01	.29	.11	.01	.28	.11	.01	.28	.11	.01
R^2 Student achievement T2	.90	.05	<.001	.92	.03	<.001	.90	.04	<.001	.93	.03	<.001

Note. MCK = mathematics content knowledge, MPCK = mathematics pedagogical content knowledge, PID = perception, interpretation, and decision-making, T = time point.

proximal a predictor is to an outcome, the stronger its predictive power tends to be, whereas the more distal the predictor is, the weaker the effect will be (Fishbein et al., 2001).

Teachers' MCK and MPCK are relatively distal to how students learn due to their different organization, for example along the disciplinary nature of mathematics for MCK or along teaching tasks for MPCK (Shulman, 1986). It is theoretical knowledge, generalizable across different classroom situations and therefore relatively abstract. Thus, models including this type of teacher cognition solely as predictors of student achievement disregard not only classroom interaction but also the situated nature of teacher cognition. The study by Baumert et al. (2010) tried to address this challenge by comparing black-box models with direct relations of teacher knowledge to students' learning progression with mediation models that included instructional quality measures. Kelcey et al. (2019) examined the influence of different facets of instructional quality as mediator of teachers' knowledge on students' learning gains as well. These are important steps, but a differentiation of teacher knowledge and skills was still lacking albeit needed to be able to estimate whether proximal teacher competence measures are superior in predicting instructional quality and students' learning progression.

The inclusion of teachers' PID skills—being more proximal to classroom practice—was supposed to bridge this gap in our approach extending previous studies. The data revealed that these skills appear to be sufficiently proximal enough to students' learning progression to predict it significantly, though the effect size was small. However, these skills also predicted instructional quality with a large effect size. Teachers' PID skills thus appear crucial to high-quality classroom management, student support, cognitive activation, and mathematics educational structuring. With this finding, our study provides evidence to an enduring discussion about how contextualized teacher competence measures should be (e.g., Shavelson, 2010). While the study by Kersting et al. (2012) made first efforts to analyze differential effects of

knowledge (MKT scores) and skills (CVA approach) predicting student learning gains, our study provides further in-depth insights into the relevance of teachers' cognitive skills as mediators in an effect chain.

Our study revealed that teacher knowledge is relevant to student progress, thus supporting the attempts made in previous studies (e.g., Baumert et al., 2010; Hill & Chin, 2018; Hill et al., 2005; Kersting et al., 2012). First, both MCK and MPCK showed significant indirect effects of medium size on the outcome variable. Second, MPCK was a strong predictor of teachers' PID skills and MCK was a strong predictor of MPCK, both relations with large effect sizes. These results are important regarding the role of MPCK in teaching and learning because most studies examining teacher effects overlooked this dimension of teacher knowledge. Applying standardized assessment is an approach that has fairly recently been developed in empirical educational research, therefore only few studies have included such a test (e.g., Baumert et al., 2010; Hill et al., 2005). With respect to MCK, these results are important in demonstrating the need for teachers to have strong content knowledge although earlier studies had not been able to establish a link to student progress. The role of MCK in teacher education has therefore long been controversial (Wu, 2011).

Our results thus supported the effect-chain model, in particular showing that teachers' knowledge can be regarded as a precondition for teachers' PID skills. As pointed out, the proximity to student learning is according to the results of our study the decisive criterion for the placement of teacher characteristics in this effect chain; a result, which we regard as the key contribution of this study, even though an experimental design was not applied.

Our study conceptualizes in addition the relation between MCK and MPCK in line with hierarchical models of cognitive abilities. Since MCK is the broader ability, it was modelled as a precondition for MPCK. The data indeed supported an interpretation of the correlation between MPCK and MCK as an effect of MCK on MPCK. It was beyond the scope of

this article, but we would like to point out that several mathematically equivalent alternatives exists to model the relation of MCK and MPCK. Given our definition of overlapping content, namely that the MPCK items measure both some MCK and specific pedagogical content knowledge, the relation could be represented by a bifactor model (Blömeke et al., 2016). A competing model could hypothesize concurrent effects of MCK and MPCK on teachers' PID skills (see Fig. 2). Since the models are mathematically equivalent (Kline, 2016), we cannot decide empirically – based on their fit to the data – which of these approaches are correct.

Our results were mixed with respect to the relation of instructional quality to students' learning progression. A significant relation was found in the full mediation model not controlling for school type and not including a direct effect of teachers' PID skills on student progress. The effect disappeared when either one was introduced. These results revealed that the final step of the effect chain was affected by the differences between academic and non-academic tracks in the German school system, indicated by the already substantial decrease in effect size regarding the relation of student achievement between the two measurement time points when school type was controlled for in the basic models. In Germany, secondary school types—in particular the differentiation into academic versus non-academic track—constitute differential learning environments due to institutional differences (e.g., different aspiration level) and differences in students' social background composition (Maaz et al., 2008). That means that, relatively independent from the single teacher, his or her competence, and the instructional quality implemented in the classroom, just controlling for school type will explain a substantial portion of differential learning outcomes in student assessments. This apparently reduces the amount of variability in student outcomes, which could otherwise be explained by variables that are conceptually more relevant or more proximal for students' learning processes.

Relating this phenomenon to our study, the following dilemma becomes apparent: If we do not control for school type, we risk to progressively overestimate the effects of other predictors such as instructional quality on student' learning progression. If we control for school type, we risk to conservatively underestimate the effects of those other predictors. As a consequence, we look at both analysis approaches to draw appropriate conclusions. These results are in line with the study by Kelcey et al. (2019), in which the district of the school had significant influence on the relation between instructional quality and students' learning progression, making it thereby more difficult to detect the hypothesized effects.

The results revealed, on the other hand, that teachers' PID skills and instructional quality were strongly related (standardized between-level estimate = .52; see electronic supplement, Table 2), and that neither

contributed more to explaining students' learning progress than the other. This is a remarkable result as it is typical for measurement that intercorrelations among constructs measured by similar types of assessments and data sources are higher than between constructs that use different kinds of assessments and data sources (in our case, externally rated observation protocol data versus video-based teacher assessment versus paper-pencil student assessment). The differences often contribute variance to scores beyond the attribute of interest (Eid et al., 2003).

Teachers' PID skills were also the only competence facet that had a direct effect on student' progression, most likely reflecting effects of third variables not included in our model. The mediation model is the most convincing model in a conceptual sense, and our empirical findings largely support this model. However, we cannot claim to have assessed and controlled for all variables potentially relevant in our context. This was impossible for practical reasons (sample size, testing time, funding available) but given the state of research also for theoretical reasons (lack of studies that examine other variables).

8. Limitations

Although we carefully derived our hypotheses based on existing theoretical and empirical research, the terms' direct and indirect effects and our mediator analyses may suggest causal relations. However, we could only estimate correlations. While we could control for potential selection effects by including the type of school students were attending and their language background as control variables, students were neither randomly assigned to teachers nor were teacher measures taken sequentially (Pearl, 2009; VanderWeele & Vansteelandt, 2009). This restricts the possibility of causally interpreting the results. Moreover, the temporal order of the assessments was not completely in line with the theoretical model that suggests a sequential effect chain while we assessed knowledge and skills at the same time.

The weakest part of the effect chain concerned the relation of instructional quality to students' learning progression. It is generally difficult to provide findings related to development given that prior knowledge typically explains much variation in the dependent variable. This phenomenon is even more visible on the between level than on the within level. In our study, students' learning progression was additionally assessed using different instruments at the first and second measurement time point. Although all tests are based on the national standards for mathematics education in Germany and share common items, they were operationalized differently. Therefore, we did not compare absolute achievement levels but estimated relations within a multi-level framework where students' achievement is connected to their teachers (i.e., within the same state). Such relations are typically
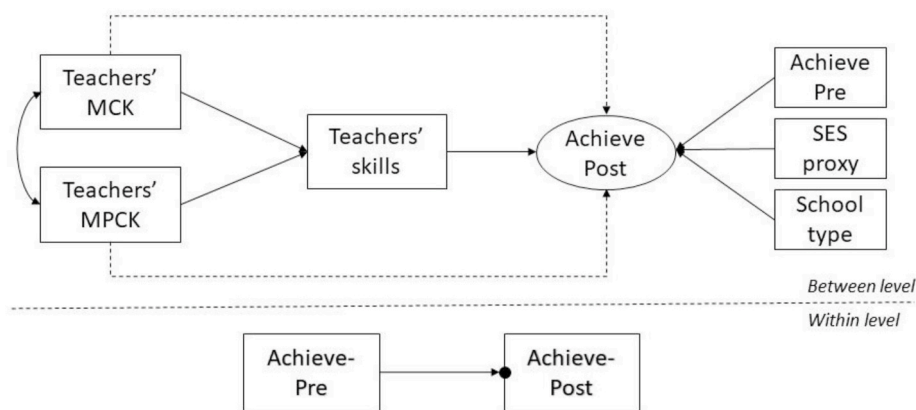
Fig. 2. Alternative model of the relation between teachers' domain-specific knowledge, PID skills, and students' learning progression in mathematics with concurrent effects of MCK and MPCK, controlling for school type and student background (dotted lines: competing hypothesis of potential direct effects of MCK and MPCK).

less prone to bias.

Another limitation is that only instructional quality could be specified as a latent variable, whereas all other constructs had to be specified as manifest variables due to limited sample size. This may have weakened the potential relations. Furthermore, since our study used data gathered in the context of educational monitoring, students' background data did not fully meet our needs. Instead of individual data, we used class composition data as a proxy measure, although such aggregation may alter a variable's meaning (Raudenbush & Bryk, 2002) and, in our context, may indicate district rather than family effects.

Finally, we must acknowledge limitations with respect to our sample. First, the teacher sample was a convenience sample, meaning we should assume self-selection bias. Second, one purpose in compiling two sub-samples from different federal states was to reach a decent sample size that allowed for testing our complex models. They would ideally have been separated to test the models' robustness across different groups. However, this would have reduced the number of units on the second level.

9. Conclusions

Overall, our results again revealed that 'black-box' models seeking to directly relate teacher knowledge to student achievement do not work well. To identify the role of teacher knowledge, it is necessary to model the relation between the different facets of teacher competence and between these and students' learning progression in a more elaborated way. Otherwise, we risk drawing conclusions based on simplified assumptions, which could potentially harm teacher education that delivers teacher knowledge (Darling-Hammond, 2006). It is therefore recommended that future studies consider effect chains including the indirect rather than direct effects of teacher knowledge.

In this context, an intriguing conceptual and methodological research question would be, which time intervals would be appropriate to assess each of the knowledge, skills, instructional quality, and student learning variables in the model so that the study design is in line with the theoretically assumed sequence. This applies especially for measures with limited stability, such as instructional quality. Some scholars have addressed this issue similarly to the present study, namely by assessing instructional quality in several lessons with adequate generalizability to explain students' gain in achievement over an extended time period (e. g., Malmberg, Hagger, Burn, Mutton, & Colls, 2010; Meyer et al., 2011). Another approach— sometimes taken in subject-specific studies—would be to assess instructional quality depending on the content that is taught in class (e.g., the Pythagorean Theorem by Lipowsky et al., 2009). However, in this case also the achievement test would have to be adjusted to the corresponding content areas ("instructional sensitivity", see e.g., Naumann et al., 2016) and would limit the generalizability of the results.

Furthermore, it is important to think about additional potential mediators because one can argue that there is still limited validity in this respect. For instance, our study design was unable to address explicit adaptations teachers may make either during lesson preparation or on the fly. Teachers need to adjust their instructional practices to students' needs and context characteristics (Klieme, 2013; Parsons et al., 2018). It is difficult to examine such non-linear relations but to learn more about adaptive processes or additional mediators on the student side (e.g., meeting the basic needs autonomy, competence, and relatedness; Ryan & Deci, 2000) would be important for being able to design teacher education (and follow-up professional development) properly.

Our results stress the importance of teachers' PID skills, which seem closest to occurrences in the classroom and to students' learning progression. Teacher education and professional development should therefore pay particular attention to developing this facet of teacher competence. We assume that PID skills do not develop naturally but must be supported by special activities involving, for example, classroom videos that train teachers' PID skills (Santagata et al., 2021; Star &

Strickland, 2008).

The results indicate that in Germany it is of key importance, whether students attend a *Gymnasium* or another type of school. The significant school type effect points to institutional differences, for example, regarding curricula, but also student composition effects (Maaz et al., 2008), which is similar to the district effect in the US (Kelcey et al., 2019). Future studies should therefore analyze the extent to which instructional quality may be determined by institutionalized grouping and how this may influence the teacher competence–instructional quality–student progression effect chain. The present study has opened up research perspectives that should be continued by future investigations into the field beyond using school type as a mere control variable.

We suggest that future studies include further facets of teacher competence and student outcomes as our focus in both respects was on cognitive constructs. However, affective-motivational constructs should also be examined because they are not only predictors and outcomes themselves but are instrumental in supporting students' learning progression. Several studies have examined these relations (e.g., Blazar & Kraft, 2017; Kunter et al., 2013) but covered only parts of a potential effect chain.

Finally, an important open question is whether our mediation model with its implicit causal relations also means that the different facets of teacher competence should be acquired sequentially. Having strong MCK may be regarded as a precondition for being able to have strong MPCK based on our results, and strong MPCK may in turn be regarded as a precondition for being able to have strong PID skills. The question is whether opportunities to learn MCK, MPCK, and PID skills during teacher education also need to be offered in this order or whether they can or even should be taken to some extent in parallel to each other or in some integrated version to facilitate connectedness of these cognitive facets. How teacher education should be designed, whether it should adopt an integrated perspective on developing teacher competence or develop these sequentially, has been and remains controversial (Flores, 2016; Santagata & Yeh, 2016). Evidence for either model is still lacking and needs further research.

CRediT author statement

Acknowledgement

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2022.101600.

References

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Ball, D., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*, 389–407.

Baumert, J., & Kunter, M. (2006). Stichwort professionelle kompetenz von lehrkraften. *Zeitschrift für Erziehungswissenschaft, 9*, 469–520.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis, 39*(1), 146–170.

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. (2015). Editorial: Approaches to competence measurement in higher education. *Zeitschrift für Psychologie, 223*, 1–2.

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). Sage Publishers.

Blömeke, S., & Olsen, R. V. (2019). Consistency of results regarding teacher effects across subjects, school levels, outcomes and countries. *Teaching and Teacher Education, 77*, 170–182.

Blum, W., Drüke-Noe, C., Hartung, R., & Köller, O. (Eds.). (2006). *Bildungsstandards Mathematik konkret: Sekundarstufe I – Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten*. Cornelsen/Scriptor.

Bruckmaier, G., Krauss, S., Blum, W., & Leiss, D. (2016). Measuring mathematics teachers' professional competence by using video clips (COACTIV video). *ZDM Mathematics Education, 48*, 111–124.

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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education, 57*, 300–314.

Dunekacke, S., Jenßen, L., & Blömeke, S. (2015). Effects of mathematics content knowledge on pre-school teachers' performance: A video-based assessment of perception and planning abilities in informal learning situations. *International Journal of Science and Mathematics Education, 13*, 267–286.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1). *Psychological Methods, 8*(1), 38–60. Model.

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.

Fishbein, M., Triandis, H. C., Kanfer, F. H., Becker, M. H., Middlestadt, S. E., & Eichler, A. (2001). Factors influencing behavior and behavior change (pp. 3–17). In A. Baum, T. R. Revenson, & J. E. Singer (Eds.), *Handbook of health psychology*. Lawrence Erlbaum.

Flores, M. A. (2016). Teacher education curriculum. In J. Loughran, & M. L. Hamilton (Eds.), *International handbook of teacher education* (pp. 187–230). Springer.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality (Working Paper; No. 11154)*. National Bureau of Economic Research.

Hill, H. C., Ball, D. L., Blunk, M., Goffney, I. M., & Rowan, B. (2007). Validating the ecological assumption: The relationship of measurescores to classroom teaching and student learning. *Measurement: Interdisciplinary Research and Perspective, 5*, 107–118.

Hill, H. C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal, 55*, 1076–1112.

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.

Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*, 169–202.

Jensen, A. R. (1998). The g factor. *The science of mental ability*. Praeger: Westport.

Jentsch, A., Schlesinger, L., Heinrichs, H., Kaiser, G., König, J., & Blömeke, S. (2021). Erfassung der fachspezifischen Qualität von Mathematikunterricht: Faktorenstruktur und Zusammenhänge zur professionellen Kompetenz von Mathematiklehrpersonen. *Journal für Mathematik-Didaktik, 42*, 97–121.

Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M., & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers – cognitive versus situated approaches. *Educational Studies in Mathematics, 94*(2), 161–182.

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. (2020). Analyses and validation of central assessment instruments of the research program TEDS-M. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student learning in German higher education. Innovative measurement approaches and research results* (pp. 29–51). Springer VS.

Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains (=Measures of effective teaching project research paper)*. Bill & Melinda Gates Foundation.

Kelcey, B., Hill, H. C., & Chin, M. (2019). Teacher mathematical knowledge, instructional quality, and student outcomes: A multilevel quantile mediation analysis. *School Effectiveness and School Improvement, 30*, 398–431.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring useable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*, 568–589.

Klieme, E. (2013). The role of large-scale assessments in research on educational effectiveness and school development. In M. Von Davier, E. Gonzalez, & I. Kirsch (Eds.), *(2013), the role of international large-scale Assessments: Perspectives from technology, economy, and educational research*. Springer.

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). Waxmann.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review, 47*, 180–195.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, J., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*, 716–725. https://doi.org/10.1037/0022-0663.100.3.716

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.

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 36*, 143–152.

Levy, J., Brunner, M., Keller, U., & Fischbach, A. (2019). Methodological issues in value-added modeling: An international review from 26 countries. *Educational Assessment, Evaluation and Accountability, 31*, 257–287.

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.

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99–106.

Malmberg, L.-E., Hagger, H., Burn, K., Mutton, T., & Colls, H. (2010). Observed classroom quality during teacher education and two years of professional practice. *Journal of Educational Psychology, 102*(4), 916–932.

Meyer, J. P., Cash, A. H., & Mashburn, A. (2011). Occasions and the reliability of classroom observations: Alternative conceptualizations and methods of analysis. *Educational Assessment, 16*, 227–243.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide: Statistical analysis with latent variables* (8th ed.). Muthén & Muthén.

Naumann, A., Hochweber, J., & Klieme, E. (2016). A psychometric framework for the evaluation of instructional sensitivity. *Educational Assessment, 21*, 89–101.

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., … Allen, M. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research, 88*(2), 205–242.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of the three basic dimensions. *ZDM Mathematics Education, 50*, 407–426.

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*(1), 4–15.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

Rowland, T., & Ruthven, K. (2011). Introduction mathematical knowledge in teaching. In T. Rowland, & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 1–5). New York: Springer.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

Santagata, R., König, J., Scheiner, T., Nguyen, H., Adleff, A.-K., Yang, X., & Kaiser, G. (2021). Mathematics teacher learning to notice: A systematic review of studies of video-based programs. *ZDM – Mathematics Education, 53*(1), 119–134.

Santagata, R., & Yeh, C. (2016). The role of perception, interpretation, and decision making in the development of beginnings teachers' competence. *ZDM Mathematics Education, 48*, 153–165.

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*(3), 475–490.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499.

Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training, 2*(1), 82–103.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Stahnke, R., Schüler, S., & Rösken-Winter, B. (2016). Teachers' perception, interpretation, and decision-making: A systematic review of empirical mathematics education research. *ZDM – Mathematics Education, 48*, 1–27.

Star, J., & Strickland, S. (2008). Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education, 11* (2), 107–125.

Tatto, M., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. *Conceptual framework*. Amsterdam, The Netherlands: IEA.

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. In *Findings from the IEA teacher education and development study in mathematics (TEDS-M). IEA*.

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface, 2*, 457–468.

Wu, H. (2011). The mis-education of mathematics teachers. *Notices of the AMS, 58*(3), 372–384.

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