CrossMark

ORIGINAL ARTICLE

Measuring mathematics teachers' professional competence by using video clips (COACTIV video)

G. Bruckmaier[1] · S. Krauss[1] · W. Blum[2] · D. Leiss[3]

Abstract The COACTIV video study is part of the COACTIV research program in which secondary mathematics teachers whose students participated in PISA 03/04 were examined, with respect to their professional knowledge, motivational orientations, beliefs, and self-regulation. In the video study, 284 German secondary mathematics teachers were asked to specify how they would continue lessons shown in three short video clips that all ended at "educationally crucial" points. From the teachers' written responses, which were coded by two independent evaluators according to five dimensions of high-quality teaching, their "situated reaction competency" (SRC) was inferred. Results relating to differences in school type (e.g., teachers from the German academic track performing better) and the relationship of SRC to other teacher characteristics (e.g., SRC was positively related to constructivist beliefs), as well as its impact on specific aspects of instructional quality, indicated the validity of the instrument.

Keywords COACTIV · Expertise · Mathematics · Pedagogical content knowledge · Professional competence · Teacher · Video study

1 Introduction

The adaptive and versatile requirements of teaching necessitate that teachers possess complex and flexibly available competencies, specifically content knowledge, pedagogical content knowledge, diagnostic competencies, self-regulation, and motivational orientations (Kunter et al. 2013a; Blömeke and Kaiser 2014). In order empirically to investigate research questions regarding these competencies (e.g., concerning their structure, school-type differences, or their impact on student learning), one needs valid and reliable instruments with which to measure them. In the COACTIV research program (*Co*gnitive *Activ*ation in the Classroom: Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students Mathematical Literacy), a variety of such instruments were developed and implemented with a large sample of German secondary mathematics teachers (who taught the grade 9/10 students assessed by the PISA study 2003/2004; see Prenzel et al. 2004 for details of the PISA supplement 2003/04 in Germany). In the present article we report on findings regarding a COACTIV instrument, in which video clips were used in order to prompt the teachers to react appropriately to critical teaching situations (for detailed information on the COACTIV research program, see Section 4.1).

There are quite a few different approaches for measuring teacher competencies (for an overview, see Lindmeier 2013; Kunter et al. 2013a; van der Vleuten 1996). Depaepe et al. (2013), for instance, provide an overview of scholars who measure pedagogical content knowledge (PCK) in mathematics. They have analyzed about 60 empirical studies that obtained data by paper-and-pencil tests, classroom observations, interviews, meeting observations, document analyses, questionnaires, and concept maps, and conclude that different perspectives can all have their specific advantages and disadvantages.

There is still a lively debate on how to conceptualize PCK appropriately (e.g., Buchholtz et al. 2014) and, in

✉ G. Bruckmaier
  georg.bruckmaier@ur.de

1 University of Regensburg, Regensburg, Germany

2 University of Kassel, Kassel, Germany

3 Leuphana University Lüneburg, Lüneburg, Germany

Springer

doing so, whether, in fact, "mathematical knowledge in teaching is located 'in the head' of the individual teacher, or is somehow a social asset, meaningful only in the context of its application" (Rowland and Ruthven 2011, p. 3). Depaepe et al. (2013) characterize these two points of view as *cognitive perspective* and *situated perspective*. The video approach presented in this article attempts to integrate both perspectives.

2 Measuring teacher competency

2.1 The cognitive and the situated perspective

The most common approach for measuring teacher competencies, especially with respect to professional knowledge, is to use paper-and-pencil tests. According to Depaepe et al. (2013), these tests are favored by proponents of the cognitive perspective, who assume that, for instance, PCK can be measured independently of the classroom context in which it is used. Such tests enable researchers to measure competencies in a standardized way, allowing them to analyze a large number of participants with relatively little effort (e.g., Ball et al. 2008; Blömeke and Kaiser 2014; Döhrmann et al. 2012; Krauss et al. 2013). However, as Depaepe et al. (2013) note, taking a cognitive perspective when measuring PCK also entails certain risks. An "authentic feeling" of the complexities and the situated nature of classroom life might be required to trigger PCK, which then interacts with other content categories of a teacher's knowledge in real-life teaching situations (Mason 2008; Petrou and Goulding 2011). This kind of critique questions the ecological validity of paper-and-pencil tests in general (e.g., Kaiser et al. 2015).

In contrast to this, still following Depaepe's approach, proponents of the situated perspective emphasize that PCK is displayed most naturally within the context in which it is enacted, that is within teaching itself. Clearly, investigating PCK by using classroom observations is the best approach for tracking the paradigm of the situated perspective. However, it is not clear how one might infer PCK by observational data and, consequently, how the competencies of the observed teachers can be compared when they are considered within different teaching situations (Depaepe et al. 2013; Lindmeier 2013). One way to bring together the authenticity of teaching on the one hand and standardization on the other is by using short video clips. The *video-based stimulated recall* approach, which is applied in the present article, in a way merges the cognitive and the situated perspective on PCK by displaying holistic teaching situations as prompts for teachers as a means of assessing their professional competencies.

2.2 Using video-clips for measuring teacher competencies

There is a long tradition of applying video techniques in teacher education, for instance, to allow teachers to reflect on their own lessons or for demonstrating best practice (Borko et al. 2011; Brophy 2004; Kuntze 2006). Their application in measuring professional competence, however, has attracted growing interest only in the last decade (e.g., Kersting et al. 2012; Knievel et al. 2015; König et al. 2014; Stürmer and Seidel 2015). In the context of competence assessment, there are various ways to implement video techniques. For example, it is possible to let teachers comment after having watched videos of their own or colleagues' lessons (van Es and Sherin 2010). With their own lessons, the video stimuli obviously vary for different teachers, meaning that the possible gain with respect to emotional involvement is at the expense of standardization (Kleinknecht and Schneider 2013). In many studies, therefore, each teacher is shown the same prerecorded scenes (real-life or reenacted), which typically focus on relevant periods (up to a few minutes) of teaching, that in most studies automatically stop at crucial moments (e.g., Lindmeier 2011). Usually, teachers either have to formulate an "educationally appropriate" continuation of the lesson (open-answer format; e.g., Lindmeier 2011), or they should choose one of several suggested options (closed-answer format). Alternatively, teachers can be asked to actively stop the movie themselves whenever they feel the need of an intervention or want to make a comment.

Up to now, several studies have been able to establish video-based stimulated recall techniques for the reliable and valid measuring of a teacher's action-oriented professional competency (e.g., Lindmeier 2011; Seidel and Stürmer 2014). In some of the studies, it could be shown that experts perform better than novices, thus confirming the discriminant validity of such instruments (cf. Seidel and Prenzel 2007). Furthermore, Kersting et al. (2012) could show the predictive validity of their video instrument for the actual discerned quality of a teacher's (real-life) instruction and, mediated by that instruction quality, for the learning gains of students. However, their results are restricted to a single topic (fractions) and based on a small sample size.

There are several suggestions for structuring the competencies measured by video techniques and for embedding such competencies in a broader theoretical framework as well. In their seminal overview on modeling competencies (not exclusively concerning the domain of teaching), Blömeke et al. (2015a) propose that situation-specific skills may mediate between "cognition" (e.g., professional knowledge as measured by paper-and-pencil tests) and "performance" (observable behavior). They

identify situation-specific skills as a holistic process that includes *perception*, *interpretation*, and *decision making* in domain-specific situations. In the video study TEDS-FU (the follow-up of the international comparative study TEDS-M; Blömeke and Kaiser 2014), Kaiser et al. (2015) apply this general competence model to teacher competencies (cf. Blömeke et al. 2015b). In their PID-model, "P" means perceiving particular events in an instructional setting, "I" stands for interpreting the perceived activities in the classroom, and "D" means decision-making, "either as anticipating a response to students' activities, or as proposing alternative instructional strategies" (Kaiser et al. 2015, p. 374). By analyzing path models that also include pedagogical content knowledge and pure content knowledge (CK) measured by paper-and pencil tests, the authors found that both of these latter knowledge categories are important for the perception and interpretation of classroom situations and for the generation of action strategies as well ("PID-competency"; Blömeke et al. 2014).

A related theoretical approach is the concept of professional vision (e.g., Goodwin 1994; Sherin et al. 2011; Seidel and Stürmer 2014), which also emphasizes the daily necessity for teachers to be able to perceive and interpret specific events in authentic classroom settings. For example, Seidel and Stürmer (2014) conceptualize professional vision with the two sub-components *noticing* and *reasoning*. According to the authors, noticing denotes whether teachers pay attention to important classroom events that are relevant for learning (i.e., goal clarity, teacher support, and learning climate), and reasoning refers to a teacher's competency to describe and explain these events and to predict possible consequences. Seidel and Stürmer found that their video instrument "Observer" provides a valid and reliable assessment of professional vision, revealing, for instance, that the number of courses on teaching and learning that a teacher has attended and his or her level of content-specific interest are closely related to higher levels of professional vision (Stürmer and Seidel 2015). It has to be noted that while PID explicitly includes the demand for "acting appropriately" (namely, the component of decision making), this requirement is missing in the operationalization of professional vision.

2.3 The situated reaction competency (SRC) and its sub-dimensions

In the present paper we analyze teachers' reactions ("situated reaction competency", or SRC) to three short video clips that are stopped automatically at educationally decisive points, with the subsequent instruction to specify how they would continue the lesson. There are several ways of analyzing and categorizing teachers' specific open answers (proposed dimensions, for instance, are "developing concepts", "connecting concepts and topics", "appropriate use of representations to explain algorithms", Kersting et al. 2012; or "elaboration", "links to evidence", "alternative strategies", Santagata et al. 2007). Although the teachers who participated in the COACTIV video-study obviously had to perceive and interpret the video stimuli (and therefore our study can theoretically be based on the approaches of PID or professional vision, respectively), only the resulting final continuation ("decision") was assessed. Thus, instead of theoretically splitting SRC into different activities (such as "perceiving and interpreting" or "explaining and predicting"), we prompted only one situation-based reaction, which was then rated according to five aspects of high-quality teaching. With respect to answer formats, Seidel and Stürmer exclusively used closed Likert scales, whereas Blömeke et al. (2014) partially elicited open answers, which, however, were guided into concrete directions (e.g., "Please phrase a task you would give to the class if you want to strengthen the students' competencies using mathematical representations", Kaiser et al. 2015, p. 380). We also used open-answer formats, but—in contrast to Kaiser et al.—we tried to grasp teachers' intuitions by allowing them to react as unrestrictedly as possible (e.g., "How would you continue this lesson?").

All teachers' answers were coded along five dimensions (student orientation, methodological accuracy, insight orientation, content-specific accuracy, and didactical opportunity), which were derived from generic aspects of high-quality teaching (both with respect to teaching mathematics and to teaching in general). For example, drawing on constructivist theories of learning (Collins et al. 2004; cf. Voss et al. 2013), teaching mathematics should be student-oriented (rather than pure lecturing) and insight-oriented (as contrasted with a mere teaching of techniques). From this perspective, the primary task of classroom instruction is to initiate and support an insightful learning process by stimulating students to actively and independently construct new knowledge that builds on their prior knowledge (Collins et al. 2004). Thus, learning environments should encourage learners to reflect deeply and engage actively with lesson content. Furthermore, it is important that learning not merely consists of acquiring factual knowledge, but also of the cumulative connection of diverse concepts and schemata. This type of learning entails the differentiation, modification, expansion, interlinking, restructuring, or rebuilding of existing knowledge structures (Cobb 1994; Greeno 1998). Moreover, the results of PISA 2000 renewed the call for more student and insight orientation in mathematics classrooms (Baumert et al. 2001). In sum, following a constructivist view, we think that these two orientations are quality dimensions, and they were coded by trained raters in our study (namely, Dimension 1 and Dimension 3; see Sect. 4.3 for details).

In addition, teachers ought to choose and apply teaching methods reflectively (e.g., Meyer 1989; Kunter et al. 2013a, p. 208 ff.). To make productive use of instructional time, teachers need to have a repertoire of various teaching methods and know how to implement and orchestrate them effectively in the classroom (Doyle 2006). In order to succeed in structuring learning processes effectively, teachers need to know how to organize and manage the classroom and keep groups of students on task. An important aspect of this potential success, with respect to teaching methods, is the reflective use of different forms of social interaction (i.e., teacher-led discussion, group work, partner work, seatwork, individual work plans, etc.; Doyle 2006). Therefore, the reflective application of teaching methods, such as choosing the appropriate social forms of learning, was another quality dimension that was rated in this study (Dimension 2).

Obviously, the mathematical content has to be explained in detail, and teaching should be a demanding orchestration of the mathematical subject matter (Baumert and Kunter 2013; Blum 2011). As early as 1986, Shulman drew attention to the importance of giving due consideration to the subject matter in studies on teaching and learning:

> In their necessary simplification of the complexities of classroom teaching, investigators ignored one central aspect of classroom life: the subject matter. The omission also characterized most other research paradigms in the study of teaching. Occasionally subject matter entered into the research as a context variable—a control characteristic for subdividing data sets by content categories […] But no one focused on the subject matter content itself […] Why this sharp distinction between content and pedagogical process? (Shulman 1986, p. 6)

Not only do teachers need to be able to cope with mathematically challenging instructional situations—and here we draw on Shulman (1986) once again—but they need a solid base of subject matter knowledge even more in order to be able to present arguments, establish connections, and thus develop students' conceptual knowledge in a way that is aligned with the typical processes of knowledge construction in mathematics. Therefore, the precise and rigorous use of mathematical concepts and procedures in teachers' answers was another quality dimension (Dimension 4).

Finally, in teaching mathematics, "didactical opportunities" may appear that are especially suited for fruitful learning processes (Biza et al. 2007; Wagenschein 1991). For example, sometimes it is possible to induce a certain cognitive conflict (Vosniadou and Verschaffel 2004), to illustrate an overarching mathematical idea (OECD 2003), or to confront students with typical misunderstandings (Smith et al. 1994; Kunter and Voss 2013). Such didactical opportunities (Dimension 5) for fruitful learning can unexpectedly appear in certain situations, or they can be provoked in a specific teaching situation (cf. Biza et al. 2007).

In sum, the following generic competency dimensions are, in our view, constitutive for teachers' SRC: (1) student orientation, (2) methodological accuracy, (3) insight orientation, (4) content-specific accuracy, and (5) didactical opportunity (the particular coding procedures and the specification of the labels will be explained in Sect. 4.3). While student orientation and the reflected application of teaching methods refer to competency aspects that are independent of a specific subject, the three other aspects are conceptually strongly subject-dependent (and thus, conceptually, close to PCK and CK). Consequently, we summarize the first two dimensions as "methodological sub-competency" (MC) and the latter three dimensions as "subject-specific sub-competency" (SC). Therefore, the overall construct of situated reaction competency (based on all five dimensions) postulated in our study is theoretically divided into two sub-competencies (see also Table 1).

2.4 Other competency aspects assessed by the COACTIV study

While the video paradigm of COACTIV is being published for the first time in this paper, results concerning other

Table 1 Postulated competencies and corresponding dimensions of SRC captured by the coding scheme

(Sub-) competencies	Dimension	Meaning of the dimension (in short)		
Situated reaction competency (SRC)				
Methodological sub-competency (MC)	Student orientation	Who is the focus of the teaching—the teacher himself/herself or the students?		
	Methodological accuracy	How precisely is the methodological procedure (e.g., social forms) depicted?		
Subject-specific sub-competency (SC)	Insight orientation	Which competency is mainly addressed—calculation skills or profound mathematical understanding?		
	Content-specific accuracy	How precisely and correctly is the content-related procedure depicted?		
	Didactical opportunity	In Video 1: Let the students realize the mistake on their own	In Video 2: Take into account several solution strategies	In Video 3: Recognize potential for mathematical modeling

competency aspects are documented in various publications (see, in particular, the comprehensive report on the COACTIV research program by Kunter et al. 2013a). In the following, we shed light on only those constructs that are relevant for the research questions in the present article.

2.4.1 Pedagogical content knowledge (PCK) and content knowledge (CK)

It goes without saying that especially the paper-and-pencil tests for PCK and CK assess related constructs (Krauss et al. 2013), and that they are therefore relevant in the context of SRC. PCK was operationalized by 23 items on the knowledge of:

– Explaining and representing mathematical contents (12 items)
– Mathematics-related student cognitions (typical errors and difficulties, 7 items)
– The potential of mathematical tasks (for multiple solution paths, 4 items).

CK was operationalized by 13 items assessing deep understanding of the content of the secondary school mathematics curriculum (for details, see Krauss et al. 2013). Interestingly, there was no correlation of PCK or CK with teaching experience, but there were large school-type differences, both with respect to the level of PCK and CK (which will be expanded on below) and with respect to the (latent) correlation of both knowledge constructs (Krauss et al. 2008b). Furthermore, PCK and CK were positively related to constructivist beliefs and negatively related to transmissive beliefs (see Sect. 2.4.2). Most importantly, PCK, but not CK, turned out to be predictively valid for instructional quality (see Sect. 2.4.3) and for students' learning progress (Baumert et al. 2010).

In the present investigation, the concept of PCK in particular will be broadened, both theoretically and empirically. Consequently, it is possible to examine the relationship between PCK and CK (as measured by the paper-and-pencil tests) on the one hand, and SRC (and its sub-dimensions) on the other hand.

2.4.2 Beliefs

Voss et al. (2013) theoretically distinguish constructivist and transmissive teacher beliefs (conceivably both about the nature of mathematics and about teaching and learning mathematics). An example of constructivist beliefs would be thinking of mathematics as a process permanently leading to new discoveries (the nature of mathematics), or supporting the idea of independent and insightful discursive learning (the teaching and learning of mathematics). In contrast,

an example of transmissive beliefs would be the view that mathematics is just a toolbox (nature of mathematics), or a receptive perspective on learning, for instance, that mathematics can best be learned by careful listening (teaching and learning mathematics). Voss et al. (2013) were able to show that constructivist beliefs also have predictive validity for instructional quality and for students' learning progress, whereas transmissive beliefs have detrimental effects on both outcomes.

2.4.3 Instructional quality

In order to assess teachers' instructional quality in COACTIV, a parsimonious model with three latent dimensions, which are each represented by multiple indicators, was developed (for details see Kunter and Voss 2013). Very briefly, the potential for *cognitive activation* was assessed in terms of the cognitive quality of the mathematical tasks implemented by the teachers in class tests (e.g., the need for mathematical argumentation; cf. Kunter et al. 2013a). Class tests were chosen because they allow valid conclusions to be drawn about the intended purposes of instruction (Kunter and Voss 2013). The dimension of *classroom management* was assessed using scales from both the student (PISA) and the teacher (COACTIV) questionnaires asking, for instance, for disruption levels or time wasted. Indicators of *individual learning support* were formed by scales from the student questionnaire, tapping various aspects of the interaction between students and teachers (Kunter and Voss 2013). Analyses of structural equation models revealed that PCK (as assessed by paper-and-pencil tests) was predictively valid for cognitive activation and learning support (Kunter et al. 2013a, 2013b). Furthermore, these analyses have shown that constructivist beliefs and teacher enthusiasm can statistically predict the level of classroom management, while teacher enthusiasm and adaptive self-regulation can statistically predict the amount of student support a teacher gives. Regarding the impact of instructional quality dimensions, cognitive activation and classroom management have been shown to explain both mathematics achievement and enjoyment of mathematics.

Finally, it has to be noted that in the German school system basically three secondary tracks can be separated, which vary with respect to the amount of subject matter and the number of pedagogical/psychological aspects covered in prospective teachers' respective university educations (for details on the German school system see Cortina and Thames 2013). While teachers from the Gymnasium (the most academic of the tracks, one which enables students to study at university) usually receive extensive training in the subject matter and less training in pedagogy (or educational psychology) at university, for the teachers of the Hauptschule (the least academic and more vocational track)

the opposite is true. It has already been demonstrated that teachers from the Gymnasium, for example, outperformed teachers from the other tracks, not only in CK but also (yet less dramatically) in PCK (Krauss et al. 2013). From this viewpoint, of course, it would be interesting to discover whether similar results may hold true concerning SRC and its sub-dimensions.

3 Research questions

The main goal of the study was the validation of the COACTIV video test. Specifically, the following research questions are addressed:

1. Are the video clips that were used adequate for reliably measuring SRC (or its dimensions)?
2. Which model fits the data structure best? For instance, can the postulated subject-independent and subject-dependent sub-competency be verified empirically?
3. Are there any school track differences?
   We expect the highest level of SRC from the Gymnasium teachers, especially regarding the subject-specific sub-competency (yet we were not sure about school type differences in the methodological sub-competency).
4. What relationship of SRC with PCK and CK (both measured by the paper-and-pencil tests) and teachers' beliefs can be found?
   We expect positive correlations of SRC (especially of the subject-dependent sub-dimensions) with PCK and CK, as well as with constructivist beliefs.
5. Is SRC (or its sub-competencies) predictively valid for the instructional quality?
   While it can be assumed that the methodological sub-competency might primarily influence classroom management, the subject-dependent sub-competency might have an effect on the level of cognitive activation.

4 Method

4.1 The COACTIV framework

The main goal of the German COACTIV research program was the examination and testing of mathematics teachers whose c1asses participated in the German PISA extension 2003/04. The structural combination of both large scale studies offered a unique opportunity to collect a broad range of data about students and their teachers, including the potential to connect teacher characteristics analytically with the teachers' lesson quality and with the students' achievement (Kunter et al. 2013b).

In the following, we present the operationalization of only SRC and its dimensions. The operationalizations of the paper-and-pencil tests of PCK and CK can be found in Krauss et al. (2013); all items on beliefs are listed in Schmeisser et al. (2013), and the related theoretical background can be found in Voss et al. (2013). Finally, the measurement of instructional quality is explained in Baumert and Kunter (2013).

4.2 The COACTIV video instrument for measuring SRC (situated reaction competency)

The video study itself, which was part of a large, computer-based instrument (Kunter et al. 2013a), started with an introductory video-example followed by four predefined answers that illustrated the procedure. The first video for participants to work on (see Fig. 1) displayed an introduction of the topic "inequalities including fractions", the second video dealt with "proportionality (rule of three)", and the third video showed the beginning of a lesson on "mean values". All videos were reenacted according to authentic video scenes from German classrooms that were recorded in the framework of the TIMSS video study (cf. Hiebert et al. 2003).

During Video 1, students convert an inequality containing fractions step by step. In the last step, one student multiplies both sides with the term "x − 1" without considering the possibility that this term might become negative (entailing the switch of the greater-than/less-than sign). The didactical opportunity here was to induce a "cognitive conflict" (e.g., Vosniadou and Verschaffel 2004) by consciously letting the students identify the mistake on their own. In the course of Video 2, several solutions (correct and incorrect) for a real-world task concerning proportionality were suggested by the students. In this case, it was possible to explicitly take into account and discuss all of their suggestions (instead of ignoring them and only focusing on the correct solution), since engaging with different opinions and viewpoints and dealing with contradictions is seen as one way to foster deep understanding of a matter (Cobb 1994; Palincsar 1998; Sfard 1998). In planning the lesson regarding Video 3, the didactical opportunity was to stress the modeling character of mathematics (cf. Kaiser and Sriraman 2006; Blum 2011) by explicitly contrasting and discussing different models of the mathematical idea of "average" or "mean value". Because we could not find alternative didactical opportunities in our theoretical analyses, nor in the piloting of our instrument, nor during the coding of the main study's answers, it seems that the identified opportunities are the only relevant ones in the three provided situations.

4.3 Coding scheme for the video instrument

Teachers were asked for continuations of the lessons presented in the three video clips (open-answer format), and

Video 1: "inequalities including fractions" (1 min 21 sec)	Video 2: "proportionality (rule of three)" (1 min 59 sec)	Video 3: "mean values" (2 min 2 sec)
Information concerning the video: "The following scene is about an introductory lesson concerning inequalities including fractions. In this class inequalities (without fractions) and equations (with fractions) have already been dealt with..."	*Information concerning the video:* "The following scene is about a teaching unit on the topic 'rule of three'. The previous lesson of this class dealt with proportionality..."	*Information concerning the video:* "The presented lesson is about elementary statistics. In the previous lesson the arithmetical mean, the median, and the mode were dealt with..."
Instruction for the teacher: "How would you continue this lesson?"	*Instruction for the teacher:* "How would you handle the students' solutions, and how would you then continue the lesson?"	*Instruction for the teacher:* "Please sketch a rough teaching design according to this task (and please tell us which ideas guided your procedure)?"

Fig. 1 Overview of the implemented videos

each continuation was then rated along all five dimensions. A comprehensive coding manual was developed (see Sect. 2.3; Table 1). Each of the five dimensions was coded trichotomously (codes 0, 1, 2; see below) for all videos. Items with no response were scored 0 (treating these responses as missing did not substantially influence the results). Since three videos were used, each teacher could accomplish a maximum of 6 points per dimension, meaning that a total maximum of 30 points was possible with respect to the five dimensions involved. A high score was interpreted, then, as an indicator for high SRC.

In detail, Dimension 1 ("student orientation") measured who was the primary agent. If the students were mostly in the center of the action (e.g., "the students should discuss, explain, calculate…"), answers were coded "2", whereas answers were coded "0" if the teacher put himself/herself in the center of the action (e.g., "I explain, name, demonstrate…"). If the teacher made no such statement, or the teacher and the students were described as acting roughly equally, the answer was coded "1". Dimension 2 ("methodological accuracy") assessed to what extent the teachers' answers contained precise

information about the specific teaching methods applied. For instance, if the implementation of group or partner work was clearly described, answers were coded "2". If a method was simply named (without a clear description of implementation), the answer was coded "1"; if no method was applied, the answer was coded "0". Dimension 3 ("insight orientation") recorded whether the focus of the teachers' answers was rather on conceptual understanding or rather on formal techniques ("calculus"). If teachers' answers were clearly insight-oriented (i.e., they aimed at conceptual or procedural understanding), they were given code "2," whereas a purely formal and calculus-driven continuation of the lesson shown was given code "0". Dimension 4 ("content-specific accuracy") measured to what extent the teachers' answers contained precise information about the content dealt with. If concepts, procedures, and/or mathematical relations were precisely named and correctly applied, the answers were coded "2" (e.g., "further transform to $3 < x$, write down the solution set, try different numbers, address case differentiation, repeat rules in dealing with inequalities, approach further examples which cover the multiplication with

negative numbers"). Code "0", however, was given if the answer only contained commonplaces that would apply to nearly every mathematics lesson (e.g., "write down students' solutions on the blackboard, make an assessment, structure the panel, do exercises"). Finally, Dimension 5 consisted of the didactical opportunity, which varies in each of the three videos. Code "2" was given if the didactical opportunity was clearly captured (the corresponding inherent didactical opportunity was already described in the introduction of the videos in Sect. 4.2).

The coding scheme covered both overall coding criteria, which applied for the respective dimensions concerning all videos and detailed specific instructions for each dimension and video (entailing several examples of answers for each code). Thus, both raters, who were trained extensively for the rating procedure, could draw on a detailed coding scheme. Answers to Video 1 were coded by the two raters independently; in the event of rater disagreement, consensus was reached through discussion. As the interrater reliability (Cohen's kappa; Landis and Koch 1977) was excellent ($\kappa_1 = 0.94$, $\kappa_2 = 0.90$, $\kappa_3 = 0.95$, $\kappa_4 = 0.95$, and $\kappa_5 = 0.92$ for Dimensions 1–5, respectively), Videos 2 and 3 were then coded by only one rater for economic reasons.

4.4 Sample

The teachers participating in COACTIV taught mathematics in the 9th-grade classes sampled within the framework of PISA 03/04 in Germany. Our teacher sample can thus be considered fairly representative of German 9th-grade mathematics teachers. A total number of 284 mathematics teachers from all secondary school tracks worked with the computer-based video instrument. For several of the subsequent analyses, these 284 teachers were split into three groups. The rationale for this division lies in the structure of the German secondary school system (see Cortina and Thames 2013).

Among the 284 teachers, 95 teachers worked in the Gymnasium (academic track), 73 in the Realschule (intermediate track), and 60 in the Hauptschule (vocational track), while 56 teachers were from various other less common secondary school types. Of the participating teachers, 176 (62.0 %) were male and the average age was 48.3 years (SD = 8.9).

4.5 Administration

The computer-based questionnaire (the video paradigm was part 3 out of 5) was conducted individually in a separate room at the teacher's school (in the afternoon of the same day when his/her students had taken the PISA test in the morning) by a trained test administrator. There were no time constraints; the teachers were not allowed to use calculators. Teachers were paid 60 euros for working with
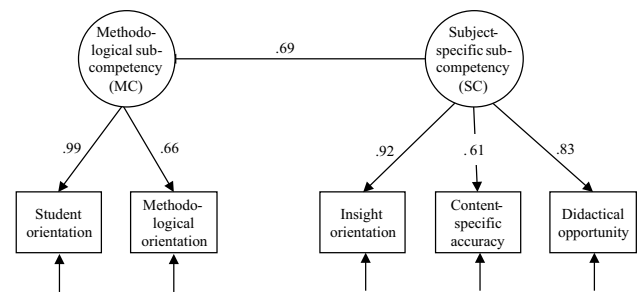
Fig. 2 2-factor measurement model of SRC

various test instruments (including the computer-based questionnaire).

4.6 Data analysis

Common statistical analyses were conducted with SPSS 21 (e.g., Field 2013); structural equation modeling was conducted with M*plus* 5.2 (Muthén and Muthén 2008). Regarding the structural equation model (Figs. 2, 3), missing values concerning aspects of instructional quality were estimated with the Full-Information-Maximum-Likelihood algorithm by M*plus*.

5 Results

5.1 Descriptives and reliability

Table 2 illustrates descriptives and reliabilities of the situated reaction competency (SRC), the two sub-competencies (MC and SC), and each single dimension. As the theoretical maximum per dimension and video is 2, arithmetic means can be considered relatively low. The reliabilities (Cronbach's α) of the five dimensions are about $\alpha = 0.50$, whereas the reliabilities of the two sub-competencies and the overall construct are substantially higher (ranging from 0.59 to 0.70). With respect to the low number of videos, the variety of topics treated in the video clips (including differing instructions), and the (entirely) open answer format, the reliabilities of SRC, MC and SC can be considered satisfactory.

Furthermore, as expected, the correlations between dimensions within the methodological sub-competency or within the subject-specific sub-competency in general were higher than between dimensions of different sub-competencies.

5.2 Dimensionality

In order to investigate the dimensionality of SRC, we applied confirmatory factor analyses. The two postulated sub-competencies were conceptualized as latent constructs based on manifest parcels (sum scores of the five
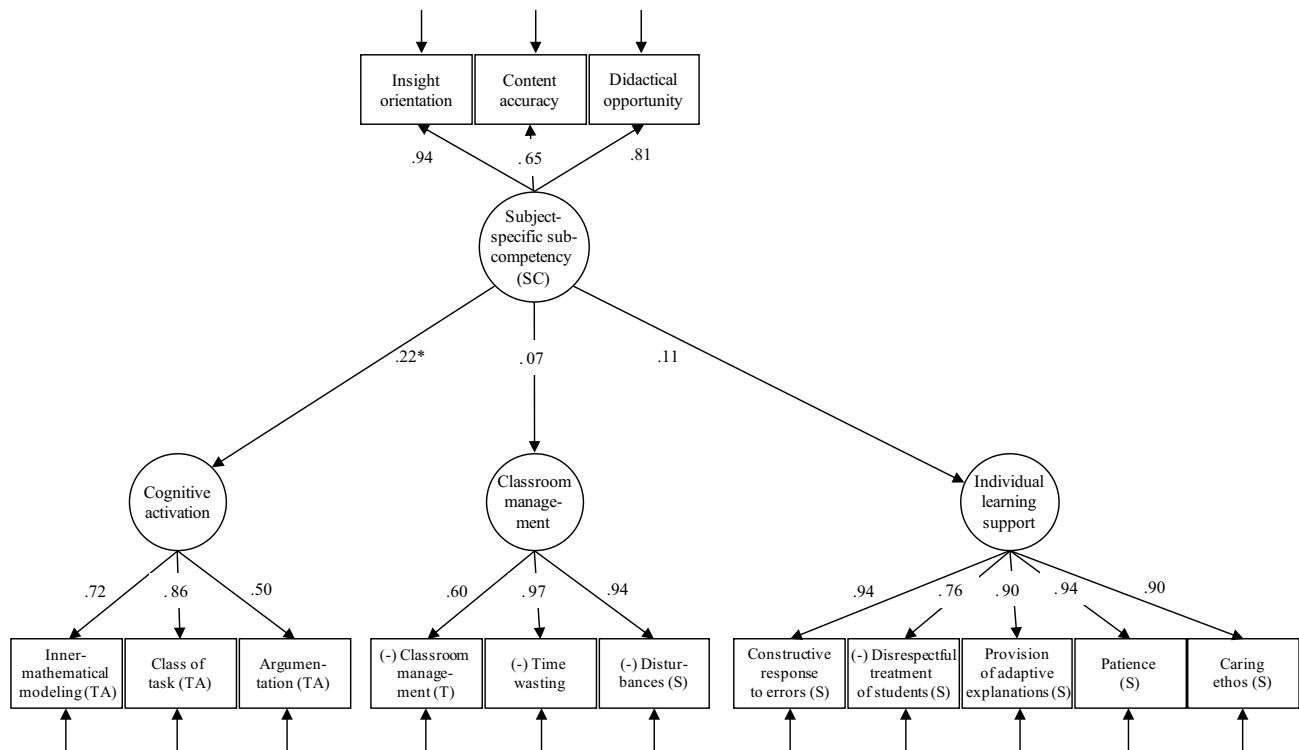
Fig. 3 Structural equation model of subject-specific sub-competency (SC) and facets of quality of teaching. The figure displays standardized loadings. Significant coefficients are indicated with a *asterisk*. (−) = scale was reverse coded, (TA) = from task analysis, (S) = from the student questionnaire, (T) = from the teacher questionnaire. Model fit: $\chi 2(p) = 117.9 (.001)$, df = 74, χ^2/df = 1.59, TLI = 0.965, CFI = 0.975, RMSEA = 0.046

Table 2 Descriptives and reliabilities of the five dimensions as well as MC, SC, and SRC

Dimensions and (sub-)competencies	Video 1	Video 2	Video 3	$\Sigma_{vid.1-3}$	Reliability
	M (SD)	M (SD)	M (SD)	M (SD)	α
Student orientation	1.30 (0.80)	1.49 (0.73)	1.30 (0.89)	4.10 (1.80)	0.59
Methodological accuracy	0.64 (0.77)	0.80 (0.77)	0.89 (0.88)	2.33 (1.72)	0.51
Insight orientation	0.65 (0.72)	1.07 (0.73)	0.84 (0.89)	2.56 (1.66)	0.51
Content-specific accuracy	0.70 (0.76)	0.71 (0.78)	0.50 (0.72)	1.91 (1.65)	0.57
Didactical opportunity	0.54 (0.77)	0.74 (0.74)	0.42 (0.64)	1.70 (1.49)	0.45
Methodological sub-competency (MC; dim. 1 + 2)	1.95 (1.39)	2.29 (1.27)	2.19 (1.66)	6.43 (3.21)	0.59
Subject-specific sub-competency (SC; dim. 3 + 4 + 5)	1.89 (1.87)	2.52 (1.74)	1.76 (1.91)	6.17 (4.17)	0.62
Situated reaction competency (SRC; dim. 1 + 2 + 3 + 4 + 5)	3.83 (2.60)	4.82 (2.55)	3.95 (3.10)	12.59 (6.50)	0.70

The theoretical maximum per dimension and video is 2, thus the maximum for the sum score per dimension is 6. Accordingly, the achievable maximum for methodological sub-competency, subject-specific sub-competency, and situated reaction competency is 12, 18, and 30, respectively

dimensions, Fig. 2). The resulting 2-factor model, however, showed non-satisfactory fit indices (Table 3).

In Table 3, various models that were based on the same covariance matrix were compared. All models based on the separation of SC from MC showed better fit indices than the competing g-factor model (while in the 2nd order model an additional superordinate latent factor was specified, in the g-factor model the five manifest sum scores were determined by one general competency). After allowing for the correlation of the residuals of dimensions 4 and 5 (as indicated by M*plus*), the modified 2-factor model showed a very good fit to the data (see Table 3), which indicates an additional underlying factor. A reason for the correlation of the residuals might be that seizing the didactical opportunity generally requires a precise description of the content involved.

Table 3 Model fit of different model specifications of SRC

Model parameters	2-Factor model	Modified 2-factor model	2nd order model	G-factor model
df	4	3	3	5
χ^2	33.54	3.66	33.54	124.27
p	0.00	0.30	0.00	0.00
RMSEA	0.16	0.03	0.19	0.29
SRMR	0.04	0.01	0.04	0.08
CFI	0.96	1.00	0.96	0.83
AIC	4807	4779	4809	4896
BIC	4865	4841	4871	4950
Gamma	0.961	0.998	0.933	–

Concerning χ^2-test, p-values > 0.05 can be considered indicative of a good model fit. RMSEA values < 0.05, SRMR < 0.08, CFI values > 0.95, and Gamma values > 0.95 indicate a good model fit. Gamma is an appropriate fit index for models with a small number of degrees of freedom, e.g., for df = 3 or 4 (Fan and Sivo 2007). Concerning AIC and BIC, low values indicate a good model fit

RMSEA root-mean-square error of approximation, *SRMR* standardized root mean residual, *CFI* comparative fit index, *AIC* Akaike's information criterion, *BIC* Bayesian information criterion

Table 4 Means M (standard deviations SD) of SRC and its dimensions by school type

Dimensions and (sub-)competencies	Theoretical range	Gymnasium M (SD) (N = 95)	Realschule M (SD) (N = 73)	Hauptschule M (SD) (N = 60)
Student orientation	0–6	4.7 (1.6)	4.4 (1.7)	3.1 (1.8)
Methodological accuracy	0–6	2.6 (1.7)	2.5 (1.8)	1.9 (1.6)
Insight orientation	0–6	3.4 (1.7)	2.4 (1.4)	1.7 (1.4)
Content-specific accuracy	0–6	2.7 (1.7)	1.8 (1.5)	0.9 (1.2)
Didactical opportunity	0–6	2.6 (1.6)	1.8 (1.3)	0.7 (0.8)
Methodological sub-competency	0–12	7.3 (3.0)	7.0 (3.1)	5.0 (3.2)
Subject-specific sub-competency	0–18	8.6 (4.2)	6.0 (3.6)	3.3 (2.6)
Situated reaction competency	0–30	15.9 (6.1)	12.9 (5.7)	8.3 (5.2)

Gymnasium: academic track; Realschule: intermediate track; Hauptschule: vocational track

5.3 School type differences

In Table 4, descriptives concerning the three most important German school tracks are displayed. Given the small numbers of COACTIV teachers in each of the other secondary school types (multi-track schools, comprehensive schools, etc.), the corresponding teacher groups will not be analyzed in detail (teachers in these other tracks showed performance levels comparable to teachers from the intermediate track).

As is evident in Table 5, there were quite substantial differences with respect to school type. These results correspond to the differences that were found in PCK and CK between Gymnasium and non-Gymnasium teachers (Krauss et al. 2013). The very large difference in the subject-specific dimensions in particular reflects the intensive coverage of mathematical subject knowledge in the Gymnasium teachers' university training, thus indicating

Table 5 Effect sizes of school type differences of SRC and its dimensions

Dimensions and (sub-) competencies	d_{G-R}	d_{R-H}	d_{G-H}
Student orientation	0.16	0.78	0.95
Methodological accuracy	0.06	0.40	0.47
Insight orientation	0.63	0.50	1.06
Content-specific accuracy	0.54	0.65	1.13
Didactical opportunity	0.55	1.00	1.43
Methodological sub-competency	0.12	0.64	0.79
Subject-specific sub-competency	0.66	0.85	1.43
Situated reaction competency	0.51	0.86	1.34

The effect size d is calculated as the mean difference divided by the pooled standard deviation. According to Cohen (1992), effect sizes of d = 0.2 can be described as small, d = 0.5 as medium, and d = 0.8 as large

G Gymnasium, *R* Realschule, *H* Hauptschule

Table 6 Correlations of SRC and its sub-competencies with teachers' professional knowledge and beliefs

Correlations	PCK	CK	Constructivist beliefs	Transmissive beliefs
Methodological sub-competency	0.20*	0.10	0.26**	−0.25**
Subject-specific sub-competency	0.36**	0.36**	0.21**	−0.31**
Situated reaction competency	0.33**	0.28**	0.26**	−0.32**

* $p < 0.05$, ** $p < 0.01$

the validity of the instrument. However, the higher performance of the Gymnasium teachers in the methodological dimensions (Dimensions 1 and 2) is nonetheless remarkable, given that these teachers usually receive less training in pedagogy (or educational psychology) at university. Consulting the open answers revealed that teachers of the Hauptschule—at least to some extent—refrained from student-oriented answers because in their teaching, they often experience students who are not capable of working independently. Thus it cannot be excluded that teachers of the Hauptschule suggested teacher-oriented continuations contrary to their theoretical knowledge of constructivist learning theories (also see Mayer 2004).

5.4 Relations of SRC to other aspects of competence

Table 6 presents the correlations of SRC and its sub-competencies with teachers' subject-specific professional knowledge (PCK and CK) and with teachers' constructivist and transmissive beliefs, all of which were measured by paper-and-pencil questionnaires in COACTIV (Krauss et al. 2013; Voss et al. 2013). These data provide first insight into the convergent and discriminant validity of the COACTIV video test. However, it has to be noted that in the correlational analyses regarding PCK and CK, teachers from the Hauptschule are not included because these teachers did not participate in COACTIV 2004 (when the PCK and CK tests were administered; in 2004 only teachers of 10th-grade classes participated, and the Hauptschule in Germany regularly ends with grade 9). Therefore, the correlational analyses regarding the PCK and CK tests are based only on N = 144 teachers. The test instruments for assessing teachers' beliefs, however, were administered in 2003.

As expected, the analyses reveal significant positive correlations of SRC both with PCK and CK (also see Blömeke et al. 2014; Kersting et al. 2012). In line with theoretical assumptions, the correlations of PCK and CK differ with respect to the two sub-competencies of SRC. While both knowledge domains correlate substantially with the subject-specific sub-competency (SC), correlations were lower with the methodological sub-competency (MC). The correlations of SC (and SRC) with pure content knowledge reflect the high interdependency of PCK and CK. This interdependency has been repeatedly verified in recent empirical studies usually interpreting CK as a necessary

prerequisite for developing PCK (Krauss et al. 2008b; Blömeke and Kaiser 2014). Accordingly, CK may also be the foundation of SC.

An interesting issue is the relationship of situated reaction competency and teachers' beliefs. Teachers with high SRC scores tended to endorse constructivist beliefs and simultaneously rejected transmissive beliefs. Differential correlations of SC and MC with the knowledge tests cannot be seen to the same extent as they are with beliefs, which can be explained by the involvement of constructivist aspects in both conceptualizations (MC: student orientation; SC: insight orientation). Together, these results indicate the convergent and discriminant validity of the COACTIV video test.

It is also interesting that no positive correlations were found between SRC or its sub-competencies and the number of years of professional experience as a teacher. These findings suggest that teachers' situated reaction competency no longer seems to develop substantially once they have completed their training. This finding seems surprising at first sight because it contradicts theories that attribute teachers' expertise development explicitly to their practical experience (Hashweh 2005; Hiebert et al. 2002; see Sect. 6 for further discussion), but it is in line with the findings mentioned in Sect. 2.4.1 concerning PCK and CK.

5.5 Predictive validity (effects on instructional quality)

Because COACTIV was linked with the PISA study (see Sect. 4.1), it was possible to relate teachers' SRC to aspects of the instructional quality as indicated by student and teacher questionnaires concerning their real-life lessons as well as by the tasks implemented by the teachers (see Sect. 2.4.3).

Two latent structural equation models were specified in order to predict the aspects of instructional quality. In one model, the methodological sub-competency was the predictor, and in the other one the subject-specific sub-competency. While analyses revealed no significant impact of the methodological sub-competency (contrary to our expectations), the subject-specific sub-competency did show a moderate to substantial influence on the aspects of instructional quality (Fig. 3).

As can be seen in Fig. 3, teachers' subject-specific sub-competency was positively related to all three aspects of

instructional quality (with only the impact on the level of cognitive activation being significant). These results indicate that insight orientation, content-specific accuracy, and the ability to seize didactical opportunities substantially influence the level of cognitive activation that teachers implement in their lessons.

6 Discussion

This paper describes the construction, implementation, and analysis of a video instrument for assessing a certain facet of mathematics teachers' professional competence. This "situated reaction competency" (SRC) was conceptualized and operationalized according to generic dimensions of high-quality teaching (student orientation, methodological accuracy, insight orientation, content-specific accuracy, and awareness of didactical opportunities). By doing so, the COACTIV video study broadens the concept of PCK as it has been pursued by COACTIV thus far. While the paper-and-pencil test of PCK can be classified into a cognitive perspective (Depaepe et al. 2013), the situated reaction competency measured in the video study is closer to a situated perspective. As was expected, the subject-specific sub-competency (SC) in particular shows substantial correlations with PCK measured by the paper-and-pencil test. Referring to the title of the present ZDM issue, this triangulation reflects the relationship between "PCK-competency" (assessed by the paper-and-pencil test) and "PCK-performance" (approximated by the video instrument).

While SRC and its sub-competencies (SC and MC) were measured reliably, the single dimensions show only restricted reliability (due to the low number of videos, the variety of topics treated in the video clips—including differing instructions—and the entirely open-answer format). Thus, only the scales SRC, MC, and SC guarantee reliable interpretation.

Several factors argue for the validity of the video instrument. First, the results regarding school track differences correspond to the expected level of expertise of the teachers examined. However, the low performance of Haupt-schule teachers may also be partially attributed to pedagogical difficulties (e.g., students' low level of autonomy) that are prevalent in this school type. Second, the correlations of SRC, SC, and MC with other characteristics of the teachers (PCK, CK, beliefs) indicate the convergent and discriminant validity of the video instrument. Regarding the paper-and-pencil PCK test, it is not straightforward to interpret the fact that also subject-matter specialists (German diploma students) show respectable performance (for details, see Krauss et al. 2008a; Krauss et al. 2013). Since the measuring of SRC is more situated, it can be hypothesized (and must be examined in further construct-validation studies) that expert teachers outperform subject-matter specialists in such a video instrument.

Interestingly, there is no positive correlation of SRC (or its sub-competencies) with work experience. However, the same result was found with respect to PCK and CK (Krauss et al. 2008a). According to deliberate practice theory, expertise does not increase simply from doing a job (Ericsson et al. 1993). Rather, motivation and deliberate practice are required to identify and overcome one's weaknesses, preferably with the support of ongoing expert feedback. Because these conditions are normally absent in everyday school life (in contrast to teacher training), our findings are in line with deliberate practice theory, the predictions of which have already been verified for various other domains (e.g., music, sports, medicine, chess).

Finally, the subject-specific sub-competency showed predictive validity regarding the level of cognitive activation. Therefore, in addition to PCK, constructivist beliefs, teaching-related enthusiasm, and aspects of diagnostic skills, another facet (SC) of teacher professionalism relevant to aspects of instructional quality could be identified in the framework of COACTIV. In contrast, facets that are not correlated to instructional quality are CK, transmissive beliefs (even negatively correlated), subject-specific enthusiasm or methodological sub-competency. As a result, what emerges is the profile of an expert teacher by identifying competence facets necessary for improving instructional quality and enhancing student learning (Kunter et al. 2013b; Sternberg and Horvath 1995; Hattie 2012). Following this path, of course, it has to be acknowledged that deviating conceptualizations and/or measurements of the involved concepts (for instance, aspects of instructional quality) might modify the results achieved so far.

While the gain of ecological validity from video-based stimulated recall approaches (as compared to paper-and-pencil tests) is obvious, they are nevertheless not optimal. It has to be noted that although video vignettes can display the context-rich reality of teaching better than text vignettes, they are still artificial for teachers (compare the two "as if" arguments by Kaiser et al. 2015). A combination of competence tests and observations of real teaching behavior might be the next step in order to diminish the gap between competence and performance (according to the title of the present issue; see also Blömeke et al. 2015a).

References

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., Klieme, E., Neubrand, J., Prenzel, M., Schiefele, U., Schneider, W., et al. (Hrsg.). (2001). *PISA 2000: Basiskompetenzen*

von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.

Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers, mathematics teacher education, 8* (pp. 25–48). New York: Springer.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Dubberke, 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. doi:10.3102/0002831209345157.

Biza, I., Nardi, E., & Zachariades, T. (2007). Using tasks to explore teacher knowledge in situation-specific contexts. *Journal of Mathematics Teacher Education, 10*(4–6), 301–309.

Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015a). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*(1), 3–13. doi:10.1027/2151-2604/a000194.

Blömeke, S., Hoth, J., Döhrmann, M., Busse, A., Kaiser, G., & König, J. (2015b). Teacher change during induction: Development of beginning primary teachers' knowledge, beliefs and performance. *International Journal of Science and Mathematics Education, 13*, 287–308.

Blömeke, S., & Kaiser, G. (2014). Theoretical framework, study design and main results of TEDS-M. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & W. H. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn* (pp. 19–48). Dordrecht: Springer.

Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J., Döhrmann, M., & Kaiser, G. (2014). Von der Lehrerausbildung in den Beruf—Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft, 17*(3), 509–542.

Blum, W. (2011). Can modelling be taught and learnt? Some answers from empirical research. In G. Kaiser, W. Blum, R. Borromeo Ferri, & G. Stillman (Eds.), *Trends in teaching and learning of mathematical modelling (ICTMA14)* (pp. 15–30). New York: Springer.

Borko, H., Koellner, K., Jacobs, J., & Seago, N. (2011). Using video representations of teaching in practice-based professional development programs. *ZDM—The International Journal on Mathematics Education, 43*(1), 175–187.

Brophy, J. (Ed.). (2004). *Using Video in Teacher Education*. Oxford: Emerald.

Buchholtz, N., Kaiser, G., & Blömeke, S. (2014). Die Erhebung mathematikdidaktischen Wissens—Konzeptualisierung einer komplexen Domäne. *Journal für Mathematikdidaktik, 35*(1), 101–128.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23*(7), 13–20.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. doi:10.1037/0033-2909.112.1.155.

Collins, A. M., Greeno, J. G., & Resnick, L. B. (2004). Educational learning theory. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 6, pp. 4276–4279). Oxford: Elsevier.

Cortina, K. S., & Thames, M. H. (2013). Teacher Education in Germany. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Eds.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers, Mathematics Teacher Education* (Vol. 8, pp. 49–62). New York: Springer.

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.

Döhrmann, M., Kaiser, G., & Blömeke, S. (2012). The conceptualisation of mathematics competencies in the international teacher education study TEDS-M. *ZDM—The International Journal on Mathematics Education, 44*(3), 325–340.

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

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*(3), 509–529.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*, 606–633. doi:10.1525/aa.1994.96.3.02a00100.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*(1), 5–26. doi:10.1037/0003-066X.53.1.5.

Hashweh, M. Z. (2005). Teacher pedagogical constructions: A reconfiguration of pedagogical content knowledge. *Teachers and Teaching: Theory and Practice, 11*(3), 273–292.

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

Hiebert, J., Gallimore, R., Garnier, H., Givven, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study*. Washington DC: U.S. Department of Education, National Center for Education Statistics.

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher, 31*(5), 3–15.

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*(2), 369–387. doi:10.1007/s10763-015-9616-7.

Kaiser, G., & Sriraman, B. (2006). A global survey of international perspectives on modelling in mathematics education. *The International Journal on Mathematics Education, 38*(3), 302–310.

Kersting, N., Givvin, K., Thompson, B., Santagata, R., & Stigler, J. (2012). Measuring usable knowledge teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568–589.

Kleinknecht, M., & Schneider, J. (2013). What do teachers think and how do they feel when they analyze videos of themselves teaching and of other teachers teaching? *Teaching and Teacher Education, 33*, 13–23.

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.

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

Krauss, S., Baumert, J., & Blum, W. (2008a). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM—The International Journal on Mathematics Education, 40*(5), 873–892.

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., et al. (2013). Mathematics teachers' domain-specific professional knowledge: Conceptualization and test construction in COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers, mathematics teacher education, 8* (pp. 147–174). New York: Springer.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., et al. (2008b). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*(3), 716–725. doi:10.1037/0022-0663.100.3.716.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2013a). *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project*. New York: Springer.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013b). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820. doi:10.1037/a0032583.

Kunter, M., & Voss, T. (2013). The model of instructional quality in COACTIV: A multicriteria analysis. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers, mathematics teacher education, 8* (pp. 97–124). New York: Springer.

Kuntze, S. (2006). Video technology in the assessment of an in-service teacher learning program. *ZDM, 38*(5), 413–421.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Lindmeier, A. (2011). *Modeling and measuring knowledge and competencies of teachers: A threefold domain-specific structure model for mathematics*. Münster: Waxmann.

Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel & K. Macha (Eds.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (pp. 45–62). Münster: Waxmann.

Mason, J. (2008). PCK and beyond. In P. Sullivan & T. Wood (Eds.), *International handbook of mathematics teacher education* (Vol. 1, pp. 301–322)., Knowledge and beliefs in mathematics teaching and teaching development Rotterdam: Sense Publishers.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*, 14–19.

Meyer, H. (1989). Plädoyer für Methodenvielfalt. *Pädagogik, 1*, 8–15.

Muthén, L. K., & Muthén, B. O. (2008). *Mplus (Version 5.1)*. Los Angeles: Muthén & Muthén.

OECD. (2003). *PISA 2003 assessment framework—mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology, 49*, 345–375. doi:10.1146/annurev.psych.49.1.345.

Petrou, M., & Goulding, M. (2011). Conceptualising teachers' mathematical knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 9–25). Dordrecht: Springer.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., et al. (Eds.). (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland—Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.

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

Santagata, R., Zannoni, C., & Stigler, J. (2007). The role of lesson analysis in preservice teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education, 10*(2), 123–140.

Schmeisser, C., Krauss, S., Bruckmaier, G., Ufer, S., & Blum, W. (2013). Transmissive and constructivist beliefs of in-service mathematics teachers and of beginning university students. In Y. Li & J. N. Moschkovich (Eds.), *Proficiency and beliefs in learning and teaching mathematics. Learning from Alan Schoenfeld and Günter Törner. Mathematics teaching and learning, 3* (pp. 51–68). Rotterdam: Sense.

Seidel, T., & Prenzel, M. (2007). Wie Lehrpersonen Unterricht wahrnehmen und einschätzen—Erfassung pädagogisch-psychologischer Kompetenzen bei Lehrpersonen mit Hilfe von Videosequenzen. *Zeitschrift für Erziehungswissenschaft, Sonderheft, 8*, 201–216.

Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in pre-service teachers. *American Educational Research Journal, 51*(4), 739–771. doi:10.3102/0002831214531321.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational researcher, 27*(2), 4–13.

Sherin, M., Jacobs, V., & Philipp, R. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York: Routledge.

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

Smith, J. P, I. I. I., Disessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115–163.

Sternberg, R. J., & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher, 24*(6), 9–17.

Stürmer, K., & Seidel, T. (2015). Assessing professional vision in teacher candidates—approaches to validate the observer extended research tool. *Zeitschrift für Psychologie, 223*(1), 54–63. doi:10.1027/2151-2604/a000200.

van der Vleuten, C. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education, 1*, 41–67.

van Es, E. A., & Sherin, M. G. (2010). The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education, 13*(2), 155–176.

Vosniadou, S., & Verschaffel, L. (2004). Extending the conceptual change approach to mathematics learning and teaching. *Special Issue of Learning and Instruction, 14*(5), 445–451.

Voss, T., Kleickmann, T., Kunter, M., & Hachfeld, A. (2013). Mathematics' teachers beliefs. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers—results from the COACTIV project. Mathematics teacher education, 8* (pp. 249–272). New York: Springer.

Wagenschein, M. (1991). *Verstehen lehren. Genetisch—sokratisch—exemplarisch*. Weinheim: Beltz.