



# Subject-specific characteristics of instructional quality in mathematics education

Lena Schlesinger<sup>1</sup> · Armin Jentsch<sup>1</sup> · Gabriele Kaiser<sup>1,2</sup> · Johannes König<sup>3</sup> · Sigrid Blömeke<sup>4</sup>

Accepted: 22 January 2018 / Published online: 31 January 2018  
© FIZ Karlsruhe 2018

## Abstract

Instructional research in German-speaking countries has conceptualized teaching quality recently according to three generic dimensions, namely, classroom management, student support and cognitive activation. However, as these dimensions are mainly regarded as generic, subject-specific aspects of mathematics instruction, e.g., the mathematical depth of argumentation or the adequacy of concept introductions, are not covered in depth. Therefore, a new instrument for the analysis of instructional quality was developed, which extended this three-dimensional framework by relevant subject-specific aspects of instructional quality. In this paper, the newly developed observational protocol is applied to three videotaped mathematics lessons from the NCTE video library of Harvard University to explore strengths and weaknesses of this instrument, and to examine in more detail how the instrument works in practice. Therefore, we used a mixed-methods design to extend the quantitative observer ratings, which enable high inference, by methods from qualitative content analysis. The results suggest the conclusion that the framework differentiates well between the lessons under a subject-specific perspective.

**Keywords** Instructional quality · Subject-specific aspects · High-inference ratings · Standardized classroom observations · Mathematics education

## 1 Introduction

The question of what good instruction is, has been intensively discussed in different fields of research for decades. Various studies could provide evidence that teachers' performance in the classroom has great impact on student achievement (Hill et al. 2005; Kersting et al. 2012; Lipowsky et al. 2009; Seidel and Shavelson 2007). At the same time, a more content-related issue arose, namely, whether and how instructional quality has to be conceptualized separately for different subjects (e.g., Klieme and Rakoczy 2008; Pianta

and Hamre 2009). In recent years, several theoretical frameworks and instruments have been developed not merely for mathematics instruction to assess different aspects of instructional quality, some of which may be called *generic* and others *domain- or subject-specific*, i.e., they are distinguished by mainly generic or subject-specific aspects of instructional quality. Therefore, they differ strongly in what they focus on when assessing instructional quality. Within recent years, the claim for the benefits of a combination of generic and subject-specific assessments has been put forward (Charalambous and Praetorius 2018).

Based on a framework widely accepted especially in German-speaking countries, namely the generic framework of three basic dimensions, we present the observational instrument developed within the study TEDS-Instruct in Germany, in which we extended the generic dimensions of instructional quality by additional subject-specific dimensions. As our previous studies and analyses mainly focused on psychometric issues such as rater agreement and the dimensionality of the construct, in this paper we apply the newly developed observational protocol to three videotaped mathematics lessons from the NCTE video library of Harvard University in order to explore strengths and weaknesses of this instrument

✉ Lena Schlesinger  
lena.schlesinger@uni-hamburg.de

Armin Jentsch  
armin.jentsch@uni-hamburg.de

<sup>1</sup> Faculty of Education, Mathematics Education, Universität Hamburg, Von-Melle-Park 8, 20146 Hamburg, Germany

<sup>2</sup> Australian Catholic University, Sydney, Australia

<sup>3</sup> Universität zu Köln, Cologne, Germany

<sup>4</sup> Centre of Educational Measurement (CEMO) at University of Oslo, Oslo, Norway

and to examine in more detail how the instrument works in practice. Furthermore, this is the first attempt to apply a mixed-methods approach to data gathered with the newly developed observational protocol, thereby allowing deeper insights than those possible by applying only quantitative methods, as has been the practice up to now.

This article is part of the ZDM Mathematics Education issue on ‘Studying instructional quality in mathematics through different lenses’, in which several different frameworks and instruments to assess instructional quality are compared and reflected on.

## 2 Theoretical rationale of the framework

For the present observational instrument we adapted the three dimensional framework of instructional quality by Klieme and colleagues, consisting of classroom management, student support and cognitive activation (Klieme and Rakoczy 2008; Lipowsky et al. 2009). The reason for choosing this framework as the basis for our instrument was its high acceptance in German-speaking countries during the last ten years. However, this framework has recently been criticized concerning the lack of content-specific aspects of instructional quality (Drollinger-Vetter 2011; Schlesinger and Jentsch 2016; Praetorius et al. 2018). Based on a systematic literature survey, we identified subject-specific characteristics of instructional quality (Schlesinger and Jentsch 2016), which were subsumed under two additional dimensions. These two subject-specific dimensions involve domain-specific aspects that are of influence to student achievements and therefore function as elements of prognostic validation. In addition, these characteristics build on solid mathematics educational theoretical frameworks (Schlesinger et al. submitted). In the following, we elaborate in more detail the reasons for extending the already existing framework by adding two subject-specific dimensions and describe how these were conceptualized.

### 2.1 Generic dimensions of instructional quality

The framework with three basic dimensions of instructional quality includes aspects that are regarded as crucial for good instruction. Even though it was firstly developed within mathematics instruction, the three basic dimensions are mainly conceptualized as generic dimensions, so that they can be used in different school subjects, and grades. Therefore, the framework is based on general theories of teaching and learning coming mainly from educational science and psychology, not limited to one subject. In the following, we describe the basic dimensions in more detail and give reasons for their relevance for good instruction referring to descriptions and conceptualizations that Klieme et al.

(2001) established as a result of the TIMSS video studies, using high-inference observer ratings analyzed with factor analyses.

The first dimension, classroom management, focuses on different discipline practices that ensure a high-quality learning time with an appropriate atmosphere without disruptions and conflicts that hinder students’ learning processes (Brophy 2000; Kunter et al. 2007; Lipowsky et al. 2009; Praetorius et al. 2012; Taut and Rakoczy 2016). Clear rules and routines, successful interventions when disruptions occur, and a structured and well-organized lesson are evidence-based characteristics of effective classroom management (Kounin 1970). The educational factor underlying this dimension is to maximize the learning time available for students. Different studies could confirm the positive impact of effective classroom management on students’ achievement (Seidel and Shavelson 2007) as well as a positive effect on students’ motivation (Rakoczy 2006).

The second generic dimension, student support, focuses on help for individual students provided by the teacher to assist students in their personal learning processes with adaptive teacher interventions and differentiated learning opportunities (Klieme et al. 2009; Lipowsky et al. 2009; Taut and Rakoczy 2016). A positive and respectful learning climate, constructive feedback and a good relationship between students and teacher are important aspects of this dimension. The psychological factor underlying this dimension is the assumption that students are motivated when they feel self-determined, i.e., being competent, responsible for their learning process, and socially integrated (Deci and Ryan 1985). Studies could provide evidence that effective student support has a positive impact on students’ motivation and interests (Lipowsky et al. 2009).

Cognitive activation, finally, focuses on the cognitive level of the instruction provided by the teacher that challenges students to activate high-level learning processes building on their existing knowledge (Hiebert and Grouws 2007; Klieme et al. 2009; Lipowsky et al. 2009; Praetorius et al. 2014; Taut and Rakoczy 2016). This dimension builds on the constructivist assumption that learning processes cannot be “drummed” into students by the teacher from the outside but that the teacher can only provide high-level cognitive learning opportunities (Bruner 1974; Helmke 2012). Especially the third dimension has a positive impact on students’ achievement, as has been shown by different studies (e.g. Baumert et al. 2010; Lipowsky et al. 2009). Regarding the subject-specific depth of these mainly generic quality dimensions, it seems that cognitive activation is probably most closely related to subject-specific aspects of instructional quality (Klieme and Rakoczy 2008; Schlesinger and Jentsch 2016). Still the dimension can be regarded and is conceptualized as being mainly generic due to its more general, psychologically oriented characteristics.

## 2.2 Subject-specific aspects of instructional quality

According to the Learning Mathematics for Teaching Project (2011) the mathematical quality of instruction is not sufficiently assessed when only generic frameworks are used for observations in mathematics classrooms. They present examples of instruction in which the lesson is well structured, students are engaged and individually supported, but the *mathematical* quality of the lesson is lacking. Blum and colleagues (2006) equally argue for a rich ‘orchestration’ of the lesson that goes beyond generic characteristics of instructional quality. Therefore, in contrast to generic dimensions, different subject-specific frameworks were developed for mathematics education. These frameworks often built on different theories of teachers’ professional knowledge for teaching.

In more detail, such mathematics educational characteristics of instructional quality may include the usage of appropriate mathematical language and various representations, well-developed teachers’ mathematical explanations at an adequate level of rigor, appropriate examples, responses to students’ mathematical errors, mathematical sense-making activities, problem-solving, proof or modeling tasks and adequate mathematical depth (e.g., including generalizations and connections) (Hiebert et al. 2003; Klieme et al. 2009; Learning Mathematics for Teaching Project 2011; Marder and Walkington 2014; Matsumura et al. 2002; Schoenfeld 2013). This listing is not meant to be complete but contains examples of what is not yet covered within generic frameworks.

Frameworks assessing the subject-specific quality of mathematics instruction are the *Uteach Observation Protocol* (UTOP, Marder and Walkington 2014), the *Mathematics-Scan* (M-Scan, Walkowiak et al. 2014), the *Instructional Quality Assessment* (IQA, Matsumura et al. 2002), the *Elementary Mathematics Classroom Observation Form* (Thompson and Davis 2014) or the *Mathematical Quality of Instruction* (MQI, Hill et al. 2008; Learning Mathematics for Teaching Project 2011) (for an overview see Schlesinger and Jentsch 2016). Finally, there are first approaches to combining the assessment of both generic and also subject-specific characteristics in mathematics instruction (e.g. *TRU Math framework*, Schoenfeld 2013). Still, to our knowledge, such an extension has not yet been developed for the framework with three basic dimensions, which are not focusing on a specific area of mathematics (e.g. geometry, quadratic equations), as was done for example in the Pythagoras study (Klieme et al. 2009).

For such an extension, we first wanted to adopt one of the existing instruments for our classroom observations and combine them with the framework of three basic dimensions. The framework by Schoenfeld seemed to be suitable as it already combined generic and subject-specific characteristics. However, it was not easily possible to match this or other frameworks with the framework of the three basic dimensions because amongst others the TRU framework

focuses primarily on students and not on the teacher. In addition, there existed no instrument that covers process-oriented aspects of mathematics education, which play an important role in the German national standards as they are regarded as crucial for the development of deep mathematical knowledge (Blum et al. 2006). As no instrument was available analyzing in depth instructional quality from a content-related perspective covering the described aspects, a new observational protocol and extension of the framework with three basic dimensions was developed referring to already existing instruments, which were enriched and further developed.

## 3 Description of the instrument

### 3.1 Subject-specific dimensions of the instrument

Based on a systematic literature review of existing classroom observation instruments (Schlesinger and Jentsch 2016), expertise in teaching mathematics within the research group, and following the discussion about the German national standards, we developed subject-specific descriptions of instructional quality, which can be empirically divided into two subject-specific dimensions (see Table 1), one covering the ‘subject-related quality’ of instruction (focus on content matter), the other one covering the ‘teaching-related quality’ (focus on practices in mathematics instruction). This classification is also consistent with the discussion on pedagogical content knowledge (Depaepe et al. 2013) that gives rise to a conceptualization with one dimension closely oriented on subject matter and the other dimension related to instructional practices (Buchholtz et al. 2014).

The first subject-specific dimension focuses on the *subject-related* quality of instruction with regard to mathematical correctness and depth (e.g. Baumert et al. 2010; Learning Mathematics for Teaching Project 2011). This first dimension is based on the assumption that students can only learn effectively if the mathematical content is presented in a correct way and covers not only superficial mathematical aspects but meaningful and deep subject matter. Summing up, it is important that the teacher is correct in his or her mathematical language and notations (Hill et al. 2008). Furthermore, teachers’ mathematical explanations and presentations need to be mathematically precise but also more in-depth and understandable for students (Schoenfeld 2013). This holds especially when responding to students’ mathematical errors that may occur during the lesson. In addition, different mathematical competencies should be dealt with during the lesson, e.g., problem solving, modeling, or proving. In the German national standards the support of these mathematical competencies in mathematics instruction is officially required (Blum et al. 2006). These national standards were developed in Germany in 2003 for the purposes

**Table 1** Operationalization of two subject-specific dimensions of instructional quality besides the common three generic dimensions

Mathematics educational quality	
Subject-related quality	Teaching-related quality
Dealing with mathematical errors of students ( <i>The teacher analyzes students' errors and misconceptions and uses them as an opportunity to learn</i> )	Using multiple representations ( <i>The teacher uses various representations within instruction and illustrates the relation between them to facilitate students' access to the mathematical content</i> )
Teacher's mathematical correctness ( <i>The teacher is mathematical correct without making content-related mistakes and being imprecise</i> )	Deliberate practice ( <i>The focus of the lesson or parts of the lesson is on practicing and the teacher provides exercises that contain opportunities for exploring, reflection and self-differentiation</i> )
Teacher's mathematical explanations ( <i>Teachers explanations are well-structured, precise and focusing on fundamental mathematical aspects</i> )	Appropriate mathematical examples ( <i>The chosen mathematical examples focus on fundamental mathematical ideas and are suitable for the students</i> )
Mathematical depth of the lesson ( <i>The teacher provides generalizations, mathematical connections and possibilities to structure the mathematical content</i> )	Relevance of mathematics for students ( <i>The teacher addresses the relevance of the mathematical content and provides connections to students' everyday life so that they may bring in their own experience and interests</i> )
Support of mathematical competencies ( <i>The teacher provides the opportunity to deal with mathematical processes such as problem-solving, modeling or reasoning and proof</i> )	

of securing the quality of mathematics instruction and of making students' achievement in German mathematics classrooms comparable. Within these national standards, the following mathematical competencies required that students should acquire, are distinguished: competence in (1) mathematical reasoning and proof, (2) problem solving, (3) mathematical modeling, (4) using different mathematical representations, (5) using mathematical language and communication, (6) calculating by working with symbolic and formal aspects and mathematical tools (Blum et al. 2006).

The second dimension consists of subject-specific aspects that are more *teaching-related*. For instance, it is important that the mathematical content be presented in a way that is accessible for the students within the instruction (Marder and Walkington 2014; Matsumura et al. 2002). This second dimension is based on the assumption that students can only learn effectively and be motivated if the mathematical content within the lesson is accessible and interesting for them. Therefore, it is necessary that different perspectives and representations be used to support the students in their learning processes and that the examples and tasks used are appropriate for the mathematical content (Drollinger-Vetter 2011; Kersting et al. 2012). Furthermore, students need to make sense of what they are learning, for example by means of the teacher providing real-world phenomena that help students to see their relevance for students' lives.

Regarding the three basic dimensions and the subject-specific extension with two dimensions, there exist some overlaps, which seem to be unavoidable. As already described above, it is a matter of discussion whether the three basic dimensions really are generic by nature or if they also cover

some subject-specific aspects. For example the item "dealing with mathematical errors of students" is part of student support in other frameworks and instruments (Praetorius et al. 2018). Still, in these instruments the focus of the item is much more on classroom climate aspects and less on content specific teacher decisions that need some pedagogical content knowledge or diagnostic competencies to analyze students' misconceptions successfully. The same holds for some overlaps to cognitive activation, as this basic dimension is probably most closely related to subject-specific aspects of instructional quality (Klieme and Rakoczy 2008; Schlesinger and Jentsch 2016). For example, the mathematical depth of the lesson probably has connections to the cognitive level of questions and tasks, and the use of different representations can also promote students' cognitive activation but simultaneously supports students in their personal learning processes. Therefore, we expect that these different dimensions cannot be completely separated.

### 3.2 Operationalization of the instrument

In order to show how we operationalized the instrument, selected item examples are presented in Table 2 as well as a range of indicators that describe incidents that can occur in mathematics instruction, and which were evaluated in order to assess instructional quality (the whole instrument can be found in the Appendix). The observational protocol consists of 26 items that are assessed by high-inference ratings for which the presented indicators describe typical examples for each item. High-inference ratings are observer ratings that have high inference in the research results. The items

**Table 2** Example items and indicators of the observation protocol

Dimension	Items	Example item	Indicators
Classroom Management	6	Effective use of lesson time	The lesson starts and ends on time Transitions between lesson phases happen smoothly Lesson time is used for content-related instruction
Student support	7	Dealing with heterogeneity	Additional materials for subgroups of students exist Tasks address different types of students The teacher offers ad-hoc differentiation during lesson (e.g. by varying the cognitive level of questions)
Cognitive Activation	5	Challenging questions and tasks	The teacher asks cognitively appropriate questions The teacher presents challenging tasks The teacher spontaneously encourages cognitive challenges
Subject-related quality	4	Mathematical depth of the lesson	Mathematical generalizations are developed The topic is related to other mathematical topics The content is embedded in a broader mathematical structure
Teaching-related quality	4	Using multiple representations	Multiple representations are used during the lesson Relations between multiple representations are shown

are formulated independently of the specific mathematical topic taught, which makes the instrument applicable for mathematics instruction at all year groups at the secondary level. Four-point Likert scales from 1 = *Does not apply at all* through 4 = *Does fully apply* were used to assess the extent to which the different characteristics were observed.

In addition, six more low-inference categories were developed that assess which mathematical process-oriented competencies are supported within the lesson (0 = not supported; 1 = supported slightly; 2 = mathematical competence in the focus of the lesson). These items are ‘usage of adequate mathematical language’, ‘promotion of mathematical modeling’, ‘promotion of problem solving’, ‘reasoning and proof’, ‘adequate usage of calculations (symbolic and formal aspects)’ and ‘adequate usage of mathematical tools’.

A coding manual was developed for the instrument that contains detailed descriptions of the items with typical examples and guidelines for the assessment (see as one example Table 3).

### 3.3 In-vivo vs. video-based observation of mathematics instruction

Our instrument was designed to be used in-vivo without videotaping the lessons. This leads to many advantages because videotaping a larger number of mathematical lessons has become extremely difficult in Germany within recent years due to legal restrictions. In contrast, in-vivo instruments have become popular especially for assessments by governmental quality insurance institutions or superintendents of schools (Pietsch and Tosana 2008). In addition, the observers in in-vivo observations are directly in the middle of the lesson and do not see the classroom through a kind of “window” as is the case with videotaped lessons. Aspects that are not covered or not in the focus of the camera are not visible for observers that view only the videotaped lesson (Casabianca et al. 2013).

Still, in-vivo observations also have some disadvantages compared to video-based lesson ratings. Observers cannot

**Table 3** Example item and detailed descriptions from the coding manual

Teacher’s mathematical correctness			
Indicators			
<ul style="list-style-type: none"> <li>• The teacher makes no content-related mistakes</li> <li>• The teacher makes no formal mistakes</li> <li>• The teacher is mathematically precise in language</li> <li>• The teacher is mathematically precise in notation</li> </ul>			
Rating of the item			
1	2	3	4
None of the four indicators applies A correct learning process is not possible due to teachers’ content-related and formal mistakes	The teacher makes a few mistakes that partly affect the learning process because some necessary conditions are presented incorrectly	The teacher makes some mistakes or is mathematically imprecise but these errors and imprecisions do not affect the learning process	All four indicators apply The teacher makes no mistakes and is mathematically precise

stop the lesson process so they need to make decisions very quickly, and there is no possibility of watching the lessons again. In addition, the audio quality is another key difference, because if the teacher works with single students these conversations can be difficult to hear in the classroom for in-vivo observers who are sitting far away. In video-based studies, the teacher normally wears a microphone that captures all conversations with students. Still, the placement of the microphone might also have disadvantages for videotaped lessons, as other conversations can get lost (Casabianca et al. 2013). Finally, videotaped lessons can be watched at any time of the day or within a research process so that time effects can be reduced. In-vivo observations have to take place when the lesson takes place.

Until now, there exist only few studies on the effect of observation modes by comparing in-vivo observations in the classroom with ratings of videotaped lessons (Casabianca et al. 2013). Casabianca et al. (2013) explored the effects of the observation mode using the CLASS-S (Secondary) instrument in 82 algebra lessons. Their results show that the mode did not have a great influence on the teacher inferences in a classroom observed over a year. Still they could show that there were some differences in the reliability and inferences in a single lesson. The mean scores and the reliability of inferences were partly higher for in-vivo than video observations. Another research group conducted live video observations—that means seeing a video of the lesson as it happens—for reducing reactivity issues (Liang 2015). The effects of this combination of video and live observation, however, are even more difficult to predict.

### 3.4 Empirical support of the instrument

The present instrument was developed and used for the first time within the study TEDS-Instruct in which the relation between teacher competencies, instructional quality, and student achievement was analyzed. 38 teachers were assessed twice (lessons of 90 min) within their instruction. To include information about the stability and variability of instructional quality, the lessons were rated multiple times, typically after periods of about 20 min. To ensure inter-rater reliability, the observations were done in-vivo by a team of six trained raters. Two observers from the team who were randomly selected rated a lesson. After each lesson the two raters carried out an intensive debriefing about the lesson and what happened within the lesson. In this debriefing it was possible to modify a rating.

In TEDS-Instruct, the generic dimensions and the two subject-specific scales reached acceptable internal consistencies (classroom management:  $\alpha = 0.86$ ; student support:  $\alpha = 0.71$ ; cognitive activation:  $\alpha = 0.82$ ; subject-related quality:  $\alpha = 0.77$ ; teaching-related quality:  $\alpha = 0.69$ ). The inter-rater reliability of  $ICC > 0.60$  for all items after

the debriefing was acceptable (classroom management,  $0.62 < ICC \leq 0.88$ ; student support,  $0.80 < ICC \leq 0.96$ ; cognitive activation,  $0.80 < ICC \leq 0.93$ ; subject-related quality,  $0.82 < ICC \leq 0.95$ ; teaching-related quality,  $0.80 < ICC \leq 0.94$ ). A two-level confirmatory factor analysis showed acceptable fit indices for five dimensions ( $\chi^2/df = 1.52$ , CFI = 0.94, RMSEA = 0.04) (Schlesinger et al. submitted).

## 4 Research questions

Three videotaped lessons were provided from the NCTE video library of Harvard University and analyzed with the standardized observational instrument to address the following research questions:

- 1) How does the instrument work in practice apart from the study in whose context it was developed?
- 2) What are the strengths and weaknesses of the newly developed instrument?

After describing the method and analysis we used, we present ratings and results for these three lessons in detail to examine advantages as well as limitations of the instrument.

## 5 Method and analysis

The three videotaped lessons from the NCTE video library of Harvard University, which we analyzed, are described in more detail in the paper by Charalambous and Praetorius (2018). To utilize this database in the best way, we applied a mixed-methods approach by combining qualitative and quantitative approaches. Quantitative methods, on one hand, tend to ‘forget’ the individual, which is at the same time the consequence of their main advantage, *generalizability* (Kelle and Buchholtz 2015). On the other hand, qualitative methods are useful to find explanations that are often not given by quantitative analyses only (Hiebert and Grouws 2007). In the present paper, we decided to use Qualitative Content Analysis (Mayring 2015) next to the quantitative observer ratings for the purpose of having the results complement each other.

In the present quantitative analysis, two experienced observers simultaneously rated instructional quality with our standardized observational instrument. The observers are two German PhD students in mathematics education who had finalized their studies for becoming teachers in mathematics education. The first and third video (Mr. Smith and Ms. Jones) were divided into two parts which were rated separately, and the second video (Ms. Young) was divided into three parts, i.e., a rating was carried out approximately every 20 min. We present descriptive results at item level for

all five dimensions aggregated across one lesson (median of the ratings). Due to the small sample size, we did no inference statistics calculations (e.g. estimation of standard errors). As we analyzed only one lesson per teacher, our results cannot be generalized to all lessons of the specific teacher. In addition to these quantitative ratings, we explain the ratings in more detail. We selected the expanded support of mathematical processes to be analyzed by Qualitative Content Analysis including six deductive categories (mathematical language, mathematical modeling, problem solving, reasoning and proof, calculations, mathematical tools), because of their core relevance for the development of deep mathematical knowledge (see Sect. 2.2). For this purpose, we used the transcribed lessons and coded these transcriptions using the six mentioned categories, every time something occurred in the text that was relevant and

appropriate to these categories. Finally we decided for each category whether the mathematical process was not supported ('0') within the lesson, supported slightly ('1') or if the mathematical process was in the focus of the lesson ('2').

## 6 Results

### 6.1 Descriptive statistics

We first present an overview of the descriptive statistics of the video analysis (Table 4) and then explain the ratings in more detail. We would like to stress that we analyzed only one lesson per teacher, so our results cannot be generalized as comprehensive personal evaluation.

**Table 4** Median of two observers' ratings of all measurement points (two measurement points for Lesson 1 and 3 and three measurement points for Lesson 2) (1 = Does not apply at all; 4 = Does fully apply)

Item	Lesson 1 (Mr. Smith)	Lesson 2 (Ms. Young)	Lesson 3 (Ms. Jones)
<b>Classroom management</b>			
Effective use of lesson time	4	3	4
Clear rules and routines	3	3	4
Preventing disturbances	3	3	3
Productive atmosphere	3.5	3	3.5
Advance organization/structuring learning processes	2	3	2
Lesson structure	2	3	2
<b>Student support</b>			
Students' individual support	1	2	2
Dealing with heterogeneity	1	1.5	1
Self-directed learning	1	2	2
Teacher feedback to students	2	3	2.5
Teacher appreciation of students	3	2	3
Student feedback	1	1	1
Support of collaborative learning	1	2	2
<b>Cognitive activation</b>			
Challenging questions and tasks	1	3	2
Supporting metacognition	1	1.5	1
Activation of prior knowledge/ co-construction	2	3	2
Cognitively challenging teaching methods	2	2	2
Facilitating remembering and recalling	2	2.5	3
<b>Subject-related quality</b>			
Dealing with students mathematical errors	n/o	2.5	2
Teachers mathematical correctness	3.5	4	4
Teachers mathematical explanations	2	3	2
Mathematical depth of the lesson	1.5	3	1.5
<b>Teaching-related quality</b>			
Using multiple representations	2	3	1.5
Deliberate practice	1	n/o	1
Appropriate mathematical examples	2	3	2
Relevance of mathematics for students	1	2	1

n/o Not observable (because amongst other aspects the focus of the lesson is not on practicing)

### 6.1.1 Lesson by Mr. Smith

The first lesson by Mr. Smith is about 40 min long. The lesson is teacher-centered with the whole class using the smartboard. The topic of the lesson is geometry and Mr. Smith teaches his class angles.

**6.1.1.1 Classroom management** The lesson by Mr. Smith is characterized by high ratings in classroom management regarding discipline practices and learning time. The atmosphere is productive and no disruptions occur. The lesson time is used for content-related instruction, routines for the organization of the lesson are apparent even though the students did not take part in organizing the lesson. Deductions are made for the advance organization and the lesson structure because Mr. Smith does not inform the students about the lesson objectives or the structuring of the learning process and the lesson is not clearly separated into sections.

**6.1.1.2 Student support** The student support by Mr. Smith is rated low except for his appreciation of the students. He does not ask about students' individual difficulties and he does not provide individual assistance for students. Mr. Smith does not offer any differentiation during the lesson and the students have no opportunity to work in a self-directed manner. Mr. Smith's answers and feedback to students are often short without giving constructive and forward-looking feedback.

**6.1.1.3 Cognitive activation** The cognitive activation of Mr. Smith's lesson is also rated low. The questions and tasks within the lesson are not cognitively challenging and only repeat students' knowledge of angles of previous lessons. Although he activates their knowledge a few times, no co-construction and further development is visible. There are few cognitively activating teaching methods used within the lesson even though several students always answer in chorus to his questions. Mr. Smith does not provide any time for metacognitive processes and students have no opportunity to reflect on their learning process.

**6.1.1.4 Subject-related quality** As his questions focus mainly on definitions without fostering high-level thinking in his students, and they require only single-word answers by the students, no students' errors or incorrect answers occurred that he needs to deal with. In addition, he does not explore any students' misconceptions. His ratings on 'teacher's mathematical correctness' are quite high because he makes no formal mistakes. Still, he is a few times mathematically imprecise within the first 20 min ("Can an angle be greater than 360 degrees?"). Mr. Smith does not give any detailed mathematical explanations over the whole lesson and does not slow down in order to address important aspects

that need more detailed explanations for his students. The mathematical depth of the lesson is overall quite low. Mr. Smith makes almost no mathematical connections or generalizations within the whole lesson and only structures the knowledge on angles but does not address any mathematical concepts. Overall, the ratings for the subject-related quality of the lesson are mainly low.

**6.1.1.5 Teaching-related quality** Mr. Smith uses two different representations of angles (symbolic and figurative), but they are not always clearly connected to each other. The focus of the lesson is on practicing existing knowledge, but Mr. Smith does not explain the importance of the exercises and there are no opportunities for exploring or reflection. The examples that Mr. Smith uses within the lesson are often at a low aspiration level without focusing on real life problems. They do not vary over the lesson and are often repeating the theme. Still, they fit the topic dealt with and address different important angles, and therefore, he received the second lowest rating for this item. The relevance of the topic is not made clear to the students at all; they are operating with angles without addressing why angles could be relevant for them. In addition, they cannot bring in personal experiences or interests. Overall, the ratings for the teaching-related quality of the lesson are low.

### 6.1.2 Lesson by Ms. Young

The second lesson, by Ms. Young, is 70 min long and the topic is about multiplying whole numbers and doubling and halving factors. The objective of the lesson is to investigate how doubling and halving factors affect the product.

**6.1.2.1 Classroom management** The lesson by Ms. Jones is characterized by a moderate rating of classroom management. The lesson time is mainly used for content-related instruction even though a few minutes are lost due to organizational problems. Rules and routines are apparent but Ms. Young needs to repeat them more often so that it seems that they are not completely accepted and implemented in the class. Ms. Young explains precisely what will happen within the lesson and describes the objective of the lesson at the beginning. The atmosphere is productive most of the time but Ms. Young is not always aware of everything that happens in the classroom.

**6.1.2.2 Student support** The student support by Ms. Young is mainly rated as low. Her appreciation of students is low, she is often rigorous and impatient, but her feedback to the students is most of the time constructive and quite sophisticated ("Think about it. I'll come back. You know what you're trying to say, but you don't know how to communicate it. So I'll give a few minutes to think about it.") The



whole class works on the same problem, so there is nearly no differentiation within the lesson. The students work most of the time in small groups and Ms. Young walks around in the class, providing individual assistance to students a few times. Still, the lesson is very teacher-centered, i.e., Ms. Young does not use the potential of self-directed learning and collaborative learning processes between students.

**6.1.2.3 Cognitive activation** The cognitive activation of the lesson by Ms. Young is rated as moderate. The questions and tasks she provides for the students are cognitively challenging and after Ms. Young activates students' prior knowledge on the task, they develop their knowledge co-constructively within the lesson. Ms. Young partly provides time for metacognitive processes and students can reflect on their learning process ("That's why we are looking for strategies. Strategies that will help us to be—to get to our answer quickly without wasting too much time, efficient ones. So some problems we offer you the opportunity to do this."). The methods used within the lesson are mainly not cognitively challenging; at the end of the lesson only a few minutes are spent on securing the acquired knowledge within the class.

**6.1.2.4 Subject-related quality** For 'dealing with students errors' Ms. Young receives ratings between '2' and '3' because she sometimes is very fast in correcting students' errors without using mistakes or ideas as a learning opportunity or analyzing errors in more detail. In addition, tolerance for mistakes is not always visible. Her ratings on 'teachers' mathematical correctness' are high. Ms. Young's mathematical explanations are more detailed than the explanations by Mr. Smith. She explains in detail why the strategy for multiplying whole numbers is useful and how it works, but sometimes her explanations are long and not well organized so that they are not always focused on the essential aspects appropriate for the students. The mathematical depth of the lesson is higher than that of the lesson by Mr. Smith and the mathematical content is structured, generalized and abstracted from the given examples. Furthermore, the lesson offers the opportunity for the students to engage in reasoning and proving. Overall, her ratings for the subject-related quality of the lesson are mainly moderate to high.

**6.1.2.5 Teaching-related quality** Within the lesson, Ms. Young and her students use different representations of multiplying whole numbers (verbal, symbolic and figurative—the students can use pictures, arrays, cubes, or story problems) that they link to each other multiple times, therefore she receives high ratings for this item. The focus of the lesson is not on practicing, therefore the rater could not rate this item. The examples that Ms. Young uses within the lesson are appropriate to the topic dealt with and they are

partly connected to real life (apples in shelves). Still these examples are not personally relevant for the students and not connected to their personal interests or experiences, therefore Ms. Young receives the second lowest rating category for the 'relevance of mathematics'. Overall her ratings for the teaching-related quality of the lesson are between '2' and '3'.

### 6.1.3 Lesson by Ms. Jones

The third lesson, by Ms. Jones, is 56 min long. The topic is about multiplying a whole number by a fraction. The goal of the lesson is to learn three different ways for this multiplication.

**6.1.3.1 Classroom management** The lesson by Ms. Jones is characterized by a mainly high rating of classroom management. Clear rules and routines are visible and Ms. Jones has a good overview of her class ("When I see that everybody in the room has their hand on top of their head, I will know we're ready to move on"). The lesson time is used for content-related instruction and Ms. Jones prevents disturbances successfully most of the time, so that a productive atmosphere exists for nearly the whole lesson. Deductions are given for the lesson structure and organization as the lesson objectives are not really clear at the beginning (it is mainly a 'cold start' without knowing what will happen), and the lesson is not clearly separated into sections, but switches between single/group work and whole class mode unexpectedly.

**6.1.3.2 Student support** The student support by Ms. Jones is mainly rated moderate. She walks around in the class and takes some time for individual students, but most of the time this is just for the purpose of checking the progress of the students and not for giving individual assistance. Ms. Jones does not offer any differentiation during the lesson and the students have few opportunities for self-directed work. Her appreciation of students is mainly high but her feedback to students is often short and not sophisticated ("Good job, student O. Good job, student M."). At the end of the lesson she supports some collaborative learning processes but these are not in the focus of the lesson. Like Mr. Smith and Ms. Young, Ms. Jones does not ask students for feedback.

**6.1.3.3 Cognitive activation** The cognitive activation of Ms. Jones lesson is rated low to moderate. The questions she asks and tasks she presents are mainly not challenging and the students only copy down what the teacher is presenting by creating posters with different strategies of multiplying a whole number with a fraction. As in the case of Mr. Smith, Ms. Jones does not provide any time for metacognitive processes, and students have no opportu-

nity to reflect on their learning processes. Ms. Jones partly activates students' prior knowledge but the knowledge of multiplying a whole number by a fraction is not developed co-constructively in class but merely presented in very small steps by Ms. Jones ("I'm going to teach you three ways to do it.").

**6.1.3.4 Subject-related quality** Within the lesson Ms. Jones often does not use students' errors as an opportunity for further learning processes and she does not analyze or address students' wrong thinking processes to help them to understand their mistakes. In contrast, she often just waits for the right answer by the students, even if they apparently did not understand their mistakes. Therefore, her 'dealing with student errors' is only rated as '2'. Like Mr. Smith and Ms. Young, Ms. Jones also gets high ratings for her 'mathematical correctness', because she does not make any formal mistakes. Her ratings on mathematical explanations are only at the low level, because the examples she uses within her explanations do not fit the problem (patting a baby on the back five times or handing in blocks five times). In addition, her explanations do not focus on why the strategies and procedures work. The mathematical depth of the lesson is rated low, because Ms. Jones does not develop any connections to other topics or mathematical content and does not attempt any generalizations or abstractions. The whole lesson focuses on procedures ("We can't do any math if we don't have numbers"). Overall, the ratings for the subject-related quality of the lesson are mainly low ('2').

**6.1.3.5 Teaching-related quality** At the beginning of the lesson, there is only one representation used for multiplying a whole number with a fraction (symbolic), but throughout the lesson Ms. Jones offers other representations (figurative and also verbal) even though they are not always well connected to each other. The lesson is rated as '1' for 'deliberate practice', because the exercises that the students conduct are not focused on students' mathematical understanding regarding the mathematical concept, and the exercises are neither self-differentiated nor reflective, nor do they give students the opportunity to discover new mathematical content. Similarly to the lesson by Mr. Smith, the relevance of the topic is not made clear to the students at all and they cannot bring in any personal experiences or interests. The real life examples that Ms. Jones tries to include even do not fit well with the topic of the lesson and do not focus on the essential aspects (multiplying with a fraction) of the mathematical topic. Therefore, the ratings on the usage of 'appropriate examples' are low as well. Overall, the ratings for the teaching-related quality of the lesson are mainly low (between '1' and '2').

## 6.2 Qualitative analyses for the support of mathematical competencies

Regarding the 'support of mathematical competencies', the two raters coded the lessons focusing on the following six categories: usage of adequate mathematical language, promotion of mathematical modeling, promotion of problem solving, mathematical reasoning and proof, adequate usage of calculations (symbolic and formal aspects) and of mathematical tools, which reflect process-oriented mathematical competencies.

### 6.2.1 Mathematical language

Promoting mathematical language was slightly supported in the first and third lesson (see the examples below). This was particularly the case when the teachers asked students for the names of mathematical objects. The lesson by Ms. Young did not focus on promoting mathematical language.

First example (first lesson):

Mr. Smith: Okay. Point B. So, I can see both are rays and B is what starts both those rays. What is a name for where two rays or where two—we could say line segments if we drew line segments—what's the name where those two meet?

Students: Vertex.

Mr. Smith: Vertex. Okay. And when I have more than one, what's the plural of that word?

Second example (third lesson):

Ms. Jones: Very good. So I take 15 and I put inside. It becomes my dividend. And 4 becomes—what is that word that we use for the number that's outside the box? Raise your hand. What is that word that we use, student R?

Student: The divisor.

Ms. Jones: Divisor. So 15 becomes my dividend and 4 becomes my divisor, and I divide it out.

### 6.2.2 Mathematical modeling and problem solving

Modeling tasks could not be observed during the three lessons, i.e., the raters could not identify any tasks which fell under the category mathematical modeling. Regarding the inclusion of mathematical problem solving, it could be slightly observed in the second lesson by Ms. Young.

Ms. Young: ... Can somebody use the story like Miss S's apples? What is happening to explain what is happening to this problem, to show that the size of the product will double if we double

one of the factors? Yes, student D, I see your hand back there. Student J, yes?

In the first and third lesson both raters did not code any situation as solving mathematical problems, i.e., students were mostly concerned with other tasks, such as calculations.

### 6.2.3 Reasoning and proof

Reasoning and proof was dealt with in the second lesson by Ms. Young and was in the focus of the lesson. However, for the first and third lesson by Mr. Smith and Ms. Jones both observers could not identify that the teachers emphasized mathematical reasoning in their classrooms.

Ms. Young: Is it true?

Students: Yes.

Ms. Young: How can we justify that? That's where we at. I want you to do it first, and then we share. How can we justify? How can we justify that 15 times 8 is the same as 30 times 4?

### 6.2.4 Calculations and mathematical tools

The usage of calculations was coded throughout all three lessons. The observers agreed that practicing (using symbolic and formal aspects) was in the focus within the first and third lesson by Mr. Smith and Ms. Jones, especially during the third one, and less focused during the second lesson by Ms. Young.

Ms. Jones: So when you were doing multiplication, it's still just repeated addition, except this time instead of adding together 2 plus 2 plus 2, you're adding together three-fourths plus three-fourths plus three-fourths.

The usage of mathematical tools was coded minimally during the first lesson dealing with geometry.

Mr. Smith: We're gonna do some stuff. You're gonna get to work on those—measuring some angles yourself. Okay. Let's measure these. Who wants to come show us how to put the protractor on one of those angles? Student G. Don't [inaudible] the protractor at the angle. Okay. What's the measure of that angle?

To sum up, in the first and third lesson teachers' foci were on training of mathematical language, i.e., repetition of mathematical concepts, and practicing (using symbolic and formal aspects). Modeling problems or proofs were not

observable during these two lessons. However, in the second lesson, the focus was on mathematical reasoning and argumentation as well as on solving mathematical problems (see Table 5).

## 7 Strengths and limitations of the instrument

In this paper, three videos of American mathematics lessons from the NCTE video library of Harvard University were analyzed with a new standardized observational protocol developed within the German mathematics educational context. The objective was to explore strengths and weaknesses of the newly developed instrument and to examine in more detail how our instrument can be used for analyzing mathematics instruction. The qualitative analysis performed in addition to the quantitative analysis could give further insights into what happened in the classroom from a content-related perspective; especially it enabled the identification of which mathematical processes were supported within the lessons.

The presented observational protocol has advantages but also some limitations. The instrument was developed for assessing instructional quality in-vivo without using video, which allows the evaluation of mathematical lessons at a broader range, as it is getting—at least in Germany—more and more difficult to get the permission of students and their parents for videotaping. Therefore, the items and indicators were developed in such a way that rating could be done quickly within the lesson. Videotaped lessons such as those in the presented cases can be rated more easily. The observation in-vivo necessitating fast rating restricts the observable complexity. In addition, the high-inference of the items may also lead to some rater disagreement because the ratings always need some estimation by the raters.

Comparable to other observational instruments, our instrument focuses on some aspects of instructional

**Table 5** Overview of the items on 'support of mathematical competencies' after quantification (0=not supported; 1=supported slightly; 2=mathematical process in the focus of the lesson)

Item	Lesson 1 (Mr. Smith)	Lesson 2 (Ms. Young)	Lesson 3 (Ms. Jones)
Mathematical language	1	0	1
Mathematical modeling	0	0	0
Problem solving	0	1	0
Reasoning and proof	0	2	0
Calculations (using symbolic and formal aspects)	1	1	2
Mathematical tools	1	0	0

quality and does not consider other aspects. The evaluation is completely standardized, so that aspects which cannot be assessed by external raters are not covered by this instrument. This standardization does not allow the observed aspects of instructional quality to be seen as an entire “to-do-list” for good instruction, especially with reference to the depth of subject-specific aspects that are observed (Steinweg 2011). Two aspects that we first wanted to include in the instrument but finally needed to remove due to raters’ difficulties in their observation and rating were teacher’s adaptivity and the separation of learning and assessment situations in the lesson. Even though these aspects are important for good instruction, it was not possible to observe these aspects in the lesson without knowing in advance the whole lesson planning of the teacher. Another advantage—but also limitation—of the instrument is its applicability in various mathematics classrooms and different instructional settings—as done in the research described in this paper—as the items and indicators are formulated independently of specific mathematical topics. However, due to this fact that the instrument does not focus on a specific mathematical topic, it is not possible to assess the conceptual coherence of the content presented (so-called ‘elements of understanding’, see Drollinger-Vetter 2011). For this purpose, the instrument would have had to be restricted to one topic or differentiated for each topic observed. Further developments may include this differentiation.

Regarding the three analyzed US American lessons, some of the items that are developed within the German context differentiate well between the lessons whereas others seem to be less adequate. Regarding the three basic dimensions, the first and third lesson by Mr. Smith and Ms. Jones received high ratings for classroom management and low ratings for cognitive activation. The lesson by Ms. Young in contrast received higher ratings for cognitive activation. Even though none of the lessons showed high student support, the lesson by Ms. Jones was rated slightly higher than the two other lessons. Regarding the subject-specific dimensions, with the exception of mathematical correctness, especially Mr. Smith and Ms. Jones received relatively low subject-specific ratings for their lessons whereas the subject-related and teaching-related mathematics educational quality of the lesson by Ms. Young were rated between moderate and high. Overall, based on these ratings, one could assume (if this lesson would be representative for their lessons in general) that Ms. Young’s class shows higher students’ achievement and that she has the highest amount of pedagogical content knowledge. However, these external stances for validation are beyond this study and the data provided. In any case, due to sample size—namely, three teachers and one lesson per each teacher—only descriptive results could be presented, which cannot be generalized to all lessons of a teacher. Regarding the support of mathematical competencies within

these three lessons, it seems that the focus on calculations and the dominance of rules and execution of algorithms may still be a characteristic of US American mathematics teaching, at least partly, which has been described as result of the TIMSS-video study (Hiebert et al. 2003).

A further advantage of the observational protocol, apart from its easy applicability and saving of resources, is its possibility of describing the variability of instructional quality throughout the lesson as the lessons can be divided into several parts (at least two parts) and these parts can be assessed separately by two raters. This facility is not common in German research currently but is more established in the American context, and it has not yet been realized for the three basic dimensions. Based on this approach it is possible to reduce rater bias, which is caused by long observation periods, as the coding is aggregated to one rating at the end of the lesson. This is especially important since some aspects of instructional quality vary more than others during a lesson (e.g., the student support or the level of cognitive activation).

The observational protocol assesses generic aspects of instructional quality (the three basic dimensions) as well as subject-specific aspects at the same time. Thus, it is possible to analyze relations between these different aspects. First approaches combined different instruments (e.g. CLASS and MQI) to analyze these relations in more detail (Blazar et al. 2017), but including generic as well as subject-specific aspects in one instrument has the advantage that researchers do not need to switch between different instruments, which would make the research process and the observations even more difficult.

Some generic and subject-specific aspects vary almost independently of each other so that it is not possible to generalize from one dimension of instructional quality to another. Until now, we have not examined these variations in detail. Thus further analyses of our data gathered with the observational protocol, maybe even with additional samples and in different contexts, are needed in order to analyze systematically the relations between the different aspects of instructional quality. First analyses indicate that the subject-specific dimensions have the highest correlation to cognitive activation (Schlesinger et al. submitted).

In contrast to other instruments—for example the TRU framework and analysis instrument by Schoenfeld (2013)—our focus is not directly on the students and their learning processes within the lesson. The focus of our instrument is primarily on the teacher and his or her instructional approaches and behavior, even though it is often not possible to assess the quality without considering students’ interactions and reactions to the teachers’ behavior. Anyway, the lack of focus on students’ reactions and behavior could be remedied by complementing the observers’ assessment with a student questionnaire implemented after the lesson.

To summarize, our newly developed instrument offers the opportunity to present a more extensive and complete picture of instructional quality, in contrast to frameworks that analyze only generic or subject-specific aspects. As the evaluation and the assessment of mathematics instruction and its quality has become more important within recent years especially with regard to the role of teachers' professional competencies in improving teaching, our observational instrument may be helpful for researchers as well as for practitioners. The instrument has the advantage that it enables the analysis of instructional quality from a generic as well as a subject-specific perspective, in order for researchers and educators to understand better what happens in the classroom and to give feedback to teachers for their professional development.

### Appendix: Items and indicators of the Observational instrument for TEDS-Instruct

Items	Indicators
<b>D1: classroom management (six items)</b>	
Effective use of lesson time	The lesson starts and ends on time Transitions between lesson phases run smoothly The lesson time is used for content-related instruction
Clear rules and routines	Patterns for the organization of the lesson are apparent The students take part in organizing the lesson
Preventing disturbances	The teacher successfully prevents emerging disturbances immediately The teacher is aware of everything that happens in the classroom
Advance organization/structuring learning processes	The teacher informs the students about the lesson objectives The teacher's expectations are apparent Tasks are given in precise language
Productive atmosphere	The volume level is appropriate Students react to teacher's signals The students and the teacher do not interrupt each other
Lesson structure	A common thread is apparent in the lesson The lesson is separated clearly into sections The teacher ends the lesson in an appropriate manner

Items	Indicators
<b>D2: student support (seven items)</b>	
Students' individual support	The teacher asks about students' individual difficulties/individual progress The teacher takes time for individuals The teacher provides individual assistance for students
Dealing with heterogeneity	Additional materials for subgroups of students exist Tasks address different types of students The teacher offers ad-hoc differentiation during lesson (e.g. by varying the cognitive level of questions)
Self-directed learning	Students check their results independently with a sample solution The teacher encourages students to work independently Students may decide whether they would like to work in groups or not
Teacher feedback to students	The teacher's feedback is sophisticated The teacher's feedback is constructive The teacher's feedback is forward-looking
Teacher appreciation of students	The teacher is patient The teacher positively enhances the students' work The teacher encourages the students
Student feedback	The teacher asks for feedback The teacher reacts to students' feedback The teacher and students talk about issues in class
Support of collaborative learning	The teacher initiates collaborative learning processes between students The teacher provides tasks that require agreement The teacher mediates interaction processes The students help each other
<b>D3: cognitive activation (five items)</b>	
Challenging questions and tasks	The teacher asks cognitively appropriate questions The teacher presents challenging tasks The teacher spontaneously encourages cognitive challenges
Supporting metacognition	At least one metacognitive sub-process takes place The teacher provides time for metacognitive processes Students reflect on their learning processes

Items	Indicators	Items	Indicators
Activation of prior knowledge/ building on students' ideas/ co-construction	The teacher asks for students' beliefs concerning the topic The students explain the task in their own words The teacher activates and explores students' prior knowledge Knowledge is developed co-constructively in class	Appropriate mathematical examples	The examples contain fundamental mathematical ideas The teacher provides an appropriate number of examples The examples are taken from students' everyday life
Cognitively challenging teaching methods	Cognitively challenging teaching methods are used The teacher provides enough time to think about the tasks The teaching methods correspond to the content and the class	Relevance of mathematics for students	The teacher provides connections to students' everyday life The teacher addresses the relevance of content Students may bring in their own experience and interests to class
Facilitating remembering and recalling	The teacher provides enough examples and helpful reminders The teacher provides enough repetitions Relevant steps are discussed with the whole class	<b>German national standards/Support of mathematical competencies (six items)</b>	
<b>D4: subject-related quality (four items)</b>		Mathematical language	The teacher provides time for building new concepts The teacher corrects language errors in an appropriate manner The teacher initiates the adequate use of mathematical language
Dealing with mathematical errors of students	The teacher uses students' errors as opportunities to learn The teacher analyzes students' errors and misconceptions The teacher is tolerant towards students' errors Students correct their errors on their own	Mathematical modeling	The teacher encourages transition processes between the real world and maths The teacher encourages the validation of students' results The teacher encourages connections between different subjects
Teacher's mathematical correctness	The teacher does not make any content-related or formal mistakes The teacher is precise concerning mathematical language and notation	Problem solving	The teacher fosters the use of heuristic strategies The teacher poses mathematical problems The teacher encourages the development of new mathematical content
Teacher's explanations	The teacher explains slowly, especially when difficulty arises The teacher focuses on the fundamental mathematical aspects Teacher's explanations are suitable for the students Teacher's explanations are well-structured and precise	Reasoning and proof	The teacher encourages mathematical reasoning The teacher demands mathematical explanations The teacher addresses mathematical proofs
Mathematical depth of the lesson	Mathematical generalizations are developed The topic is related to other mathematical topics The content is embedded in a broader mathematical structure	Calculations (using symbolic and formal aspects)	The teacher provides time for training the use of mathematical symbols The teacher provides time for training basic skills in mathematics
<b>D5: teaching-related quality (four items)</b>		Mathematical tools	The students use mathematical tools reasonably The teacher uses mathematical tools correctly
Using multiple representations	Multiple representations are used during the lesson Relations between multiple representations are shown	Items are rated on a four-point Likert scale from 1 = <i>Does not apply at all</i> through 4 = <i>Does fully apply</i>	
Deliberate practice	The teacher explains the importance of the exercises The exercises contain opportunities for exploring and reflection The exercises are self-differentiating		

## References

- 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., Braslow, D., Charalambous, C. Y., & Hill, H. C. (2017). Attending to general and mathematics-specific dimensions of teaching: Exploring factors across two observation instruments. *Educational Assessment*, 22(2), 71–94.
- Blum, W., Driike-Noe, C., Hartung, R., & Köller, O. (2006). *Bildungsstandards Mathematik: Konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichts Anregungen, Fortbildungsideen*. Berlin: Cornelsen Scriptor.
- Brophy, J. (2000). *Teaching*. Brüssel: International Academy of Education.
- Bruner, J. (1974). *Toward a theory of instruction*. Cambridge: Harvard University Press.
- Buchholtz, N., Kaiser, G., & Blömeke, S. (2014). Die Erhebung mathematikdidaktischen Wissens—Konzeptualisierung einer komplexen Domäne. *Journal für Mathematik Didaktik*, 35(1), 101–128.
- Casabianca, J. M., McCaffrey, D. F., Gitomer, D. H., Bell, C. A., Hamre, B. K., & Pianta, R. C. (2013). Effect of observation mode on measures of secondary mathematics teaching. *Educational and Psychological Measurement*, 73(5), 757–783.
- Charalambous, C., & Praetorius, A.-K. (2018). Studying mathematics instruction through different lenses: setting the ground for understanding instructional quality more comprehensively. *ZDM Mathematics Education*, 50(3). <https://doi.org/10.1007/s11858-018-0914-8>. (this issue).
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior. Perspectives in social psychology*. New York: Plenum.
- 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.
- Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit: Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht*. Münster: Waxmann.
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Hiebert, J., Gallimore, R., Garnier, H., & Stigler, J. (2003). *Teaching mathematics in seven countries. Results from the TIMSS 1999 video study*. Washington: National Center for Education Statistics.
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte: Information Age.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, 26(4), 430–511.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.
- Kelle, U., & Buchholtz, N. (2015). The combination of qualitative and quantitative research methods in mathematics education: A “mixed methods” study on the development of the professional knowledge of teachers. In A. Bikner-Ahsbahr, C. Knipping & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 321–364). Dordrecht: Springer.
- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568–589.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster: Waxmann.
- Klieme, E., & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54, 222–237.
- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: “Aufgabenkultur” und Unterrichtsgestaltung. In BMBF (Ed.), *TIMSS—Impulse für Schule und Unterricht, Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (pp. 43–58). Bonn: BMBF.
- Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart and Winston.
- Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17(5), 494–509.
- Learning Mathematics for Teaching Project (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education*, 14, 25–47.
- Liang, J. (2015). Live video classroom observation: An effective approach to reducing reactivity in collecting observational information for teacher professional development. *Journal of Education for Teaching: International Research and Pedagogy*, 41(3), 235–253.
- 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.
- Marder, M., & Walkington, C. (2014). Classroom observation and value-added models give complementary information about quality of mathematics teaching. In T. Kane, K. Kerr & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measuring Effective Teaching project* (pp. 234–277). New York: Wiley.
- Matsumura, L. C., Garnier, H. E., Pascal, J., & Valdés, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and students achievement. *Educational Assessment*, 8, 207–229.
- Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. In A. Bikner-Ahsbahr, C. Knipping & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education. Examples of methodology and methods* (pp. 365–380). Dordrecht: Springer.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Pietsch, M., & Tosana, S. (2008). Beurteilereffekte bei der Messung von Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 11(3), 430–452.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM Mathematics Education*, 50(3). <https://doi.org/10.1007/s11858-018-0918-4>. (this issue).
- Praetorius, A.-K., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction*, 22, 387–400.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Rakoczy, K. (2006). Motivationsunterstützung im Mathematikunterricht: Zur Bedeutung von Unterrichtsmerkmalen für die Wahrnehmung der Schülerinnen und Schüler. *Zeitschrift für Pädagogik*, 52(6), 822–843.

- 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., Blömeke, S., & König, J. (under review). Messung fachspezifischer Unterrichtsqualität im Mathematikunterricht. (Manuscript submitted).
- Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM Mathematics Education*, 45(4), 607–621.
- 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.
- Steinweg, A. S. (2011). Einschätzung der Qualität von Lehr-Lernsituationen im mathematischen Anfangsunterricht—ein Vorschlag. *Journal für Mathematik Didaktik*, 32(1), 1–26.
- Taut, S., & Rakoczy, K. (2016). Observing instructional quality in the context of school evaluation. *Learning and Instruction*, 46, 45–60.
- Thompson, C. J., & Davis, S. B. (2014). Classroom observation data and instruction in primary mathematics education: improving design and rigour. *Mathematics Education Research Journal*, 26(2), 301–323.
- Walkowiak, T. A., Berry, R. Q., Meyer, J. P., Rimm-Kaufman, S. E., & Ottmar, E. R. (2014). Introducing an observational measure of standards-based mathematics teaching practices: Evidence of validity and score reliability. *Educational Studies in Mathematics*, 85(1), 109–128.