



Capturing teaching practices in language-responsive mathematics classrooms Extending the TRU framework “teaching for robust understanding” to L-TRU

Susanne Prediger¹ · Philipp Neugebauer¹

Accepted: 24 August 2020

© The Author(s) 2020

Abstract

Supporting language in mathematics classrooms requires both curriculum material that follows language-responsive design principles and teaching practices that enact these principles with high instructional quality. This paper presents the analytic framework L-TRU, which was developed to assess language-responsive teaching practices quantitatively. The L-TRU framework draws upon Schoenfeld’s teaching for robust understanding (TRU) framework by adapting its five dimensions to language-responsive classrooms: Mathematical Richness, Cognitive Demand, Equitable Access, Agency, and Use of Student Contributions. It is extended by two further dimensions, namely, Discursive Demand and Connecting Registers. The adapted and extended L-TRU rating scheme was applied to 41 video-recorded lessons of 26 teachers who all used the same language-responsive curriculum material on percentages. The qualitative insights gained from selected transcripts reveal that the dimensions indeed capture important distinctions in valid ways. The analysis of interrater reliability and correlations confirms that distinct dimensions are captured with reliability. The quantitative overview of the ratings of 497 episodes shows that in spite of the shared curriculum material, a large variety of instructional practices were enacted: Consistently high quality was found in the dimensions Cognitive Demand and Equitable Access and a medium quality in Connecting Registers. The dimensions Agency, Discursive Demand and Use of Contributions show the largest variance among teachers, with Discursive Demand separating most. These findings empirically substantiate an important research tool for quantitatively capturing teaching practices with respect to their general mathematics instruction quality and language-responsive quality.

Keywords Language-responsive classrooms · Teaching practices · Quality · Rating scheme · Discursive demand · Connecting registers

1 Introduction

Students with limited access to academic language proficiency achieve less conceptual understanding in mathematics than their more language proficient peers (Secada 1992). As a consequence, language-responsive approaches were developed, such as approaches that provide better access for language learners to mathematical understanding by integrating mathematics and language-learning opportunities (Gibbons 2002; Zahner, Velazquez, Moschkovich, Vahey, and Lara-Meloy 2012). In controlled trials, these approaches have proven effective in fostering students’ conceptual

understanding (Prediger and Wessel 2013; Prediger and Neugebauer 2020). However, different classes achieved highly different learning gains. This observation of variance on the class level calls for investigating the teaching practices by which different teachers enacted the instructional approaches in their classrooms. This paper contributes to this needed research.

Many qualitative studies have identified criteria for productive teaching practices that potentially support language learners’ mathematics learning. However, quantitative evidence for these features is still rare (see Erath, Ingram, Moschkovich, and Prediger 2021 for an overview). An analytic framework with quantitative measures for capturing quality is required that combines three domains, namely, general instructional quality, mathematics-specific instructional quality, and focus on aspects relevant for promoting language.

✉ Susanne Prediger
prediger@math.tu-dortmund.de

¹ TU Dortmund University, Dortmund, Germany

In this paper, we present such an analytic framework, drawing upon Schoenfeld's (2013) teaching for robust understanding (TRU) framework. We adapt its five dimensions for language learners and extend it with two language-related dimensions to create the language-responsive TRU Framework (L-TRU). Using this framework, this paper pursues the following, mainly methodological, question: *How can we capture the instructional quality of different teaching practices in language-responsive mathematics classrooms?*

In Sect. 2, we present the theoretical background and in Sect. 3 the adapted and extended analytic framework L-TRU. In Sect. 4, the methodological background of the video study is reported. Section 5 provides qualitative insights into differences captured by the L-TRU framework and Sect. 6 the quantitative findings on interrater reliability, connections, and distribution of their quality levels. In Sect. 7 we discuss the achievements and limitations of this paper and future research needs.

2 Theoretical background: quality dimensions in generic and mathematics-related approaches for language learners

2.1 Generic instructional quality and mathematics-related instructional quality

Research on instructional quality aims at identifying features of teaching practices that are predictive for promoting students' learning. Whereas quantitative research approaches capture the statistical connections between specific practices and measurable learning gains (e.g., Brophy 2000; Hill et al. 2008), qualitative studies identify preconditions of learning in situ such as the discursive quality of the interaction (see Erath et al. 2021 for an overview). In most of these qualitative studies, researchers have focused on selected aspects of classroom situations in depth, without relating them to measurable learning gains.

As Charalambous and Praetorius (2018) summarized in their research overview, existing analytic frameworks for quantitative approaches to quality research are still mainly unconnected. Several frameworks have captured the *general instructional quality* in generic ways (e.g., the Three Basic Dimensions of classroom management, student support, and cognitive activation, by Praetorius et al. 2018), while other frameworks have tried to capture also *mathematics-related dimensions*, for instance, by elaborating on what cognitive activation or student support mean in more detail. In particular, they have involved subject-specific aspects such as mathematical richness, mathematical correctness, dealing with multiple representations, and appropriateness of the examples (e.g., Adler and Ronda 2015; Brunner 2018; Henningsen and Stein 1997; Hill et al. 2008; Schlesinger, Jentsch, Kaiser, König, and Blömeke 2018). In their early survey article, Hiebert and Grouws (2007) summarized that mathematics-related quality must include a measure for high cognitive demand and a focus on conceptual understanding. Rather than repeating the comprehensive surveys here (Charalambous and Praetorius 2018; Hiebert and Grouws 2007), in the following sections we discuss critical aspects in relation to the framework upon which we built (Sect. 2.2) and language learners (Sect. 2.3).

2.2 The TRU framework for integrating generic and mathematics-related instructional quality aiming at conceptual understanding

Schoenfeld (2013) developed the TRU framework to describe quality dimensions for different teaching practices. The ratings that were first developed for reflection in professional development later also turned into more widely used analytic tools for research purposes (Schoenfeld 2013, 2018; Schoenfeld et al. 2018).

The TRU framework starts from the richness of mathematics (similarly to Hill et al. 2008; Schlesinger et al. 2018). From there, it unfolds students' experiences with mathematics in further dimensions. The five dimensions are presented in Fig. 1 (with the core questions from Schoenfeld 2013, p.

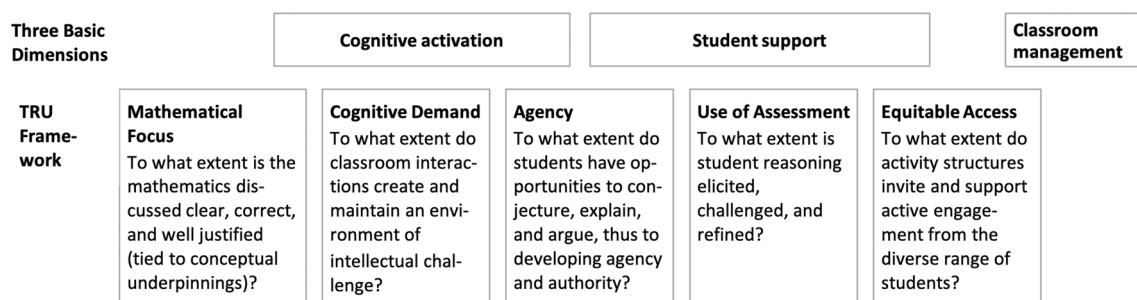


Fig. 1 TRU framework and its connection to the three basic dimensions

615/616) and related to the often cited Three Basic Dimensions (Praetorius et al. 2018).

Both frameworks overlap most with respect to the basic dimension *cognitive activation*, which is disentangled into *Cognitive Demand*, *Agency*, and *Mathematical Focus* in the TRU framework. The overlap between *cognitive activation* and *Mathematical Focus* concerns the focus on conceptual understanding, but the generic framework of Three Basic Dimensions does not cover mathematical accuracy or coherence of classroom discussions (Brunner 2018). The basic dimension *student support* is disentangled into *Agency* (the extent to which students take agency and are made accountable for their ideas) and *Uses of Assessment* (teachers' adaptivity and the degree to which they use students' reasoning for further teaching). The TRU dimension *Equitable Access* is only partly covered in *student support*, as it focuses in particular on access for underprivileged students. In contrast, the basic dimension *classroom management* has many pedagogical components that are only implicit in the TRU framework with its mathematical focus.

In each dimension, teaching practices in 5-min periods can be rated on three levels of sophistication: For example, in the dimension *Mathematical Focus*, discussions are rated as *basic* if they are “purely rote, OR disconnected or unfocused, OR consequential mistakes are left unaddressed” (Schoenfeld 2013, p. 615). They are rated as *proficient* if “mathematics discussed is relatively clear and current, BUT connections between are either cursory or lacking” (ibid.). A rating of *distinguished* is given when these connections occur. The levels for each dimension are presented in Sect. 3.

This framework is highly suitable for our focus on language learners for several reasons:

- The framework integrates generic and mathematics-related aspects in its dimensions of quality in a rather comprehensive way (Schoenfeld 2013, p. 610).
- It has been optimized with respect to high validity in many qualitative projects and to its power for professional development (Schoenfeld 2018, p. 42).
- As it emerged within the Diversity in Mathematics Education Center for Learning and Teaching (DIME 2007), it is optimized also for underprivileged students such as monolingual and multilingual language learners with its specific focus on agency and equitable access.
- The learning goal is “robust understanding”, in other words, the ability “to be effective at dealing with verbally presented, situationally based problems” (Schoenfeld 2013, p. 609), which resonates well with our focus on developing language learners' conceptual understanding (see Sect. 2.3), much more than frameworks such as the MDI framework, which was optimized for more procedural South African classroom cultures (Adler and Ronda 2015).

In spite of these advantages, the TRU framework has not yet been widely used as a tool for quantification in larger video studies (Schoenfeld 2018, p. 495). Hence, for our current usage, it was also necessary to check for interrater reliability and dimensionality of the adapted framework.

2.3 Instructional quality for language learners

The TRU framework covers many aspects formerly identified as crucial for our specific target group of (monolingual and multilingual) language learners. Following the research overview in the ZDM introduction (Erath et al. 2021), we see resonance, but there was also a need for refinement:

- *Mathematical Focus* and *Cognitive Demand*. Rich mathematics and high cognitive demands have been shown to be crucial for all students' mathematics learning (Henningesen and Stein 1997; Hiebert and Grouws 2007). However, research in the equity perspective revealed that underprivileged learners are rarely exposed to rich mathematics, and their interaction opportunities with mathematics comprise too few cognitive demands (DIME 2007; Ing et al. 2015). Possible explanations for this opportunity gap are teachers' low expectations of underprivileged learners (Callahan 2005; Jackson, Gibbons, and Sharpe 2017) who are often not engaged in sufficiently rich discourse practices relevant for enacting mathematical richness and high cognitive demand (Herbel-Eisenmann, Coppin, Wagner, and Pimm 2011). Limited academic language proficiency is particularly restricting for this access to challenging discourse practices (Moschkovich 2015), so in order to study these two explanations empirically, the discursive demand should be captured separately (see Sect. 3.2).
- *Equitable Access*. As classroom discourses are language-bound, the active participation of language learners is often limited. Thus, equitable access to rich mathematics for *all* learners is a crucial dimension for maintaining *high quality* learning opportunities (Herbel-Eisenmann et al. 2011).
- *Agency*. Strengthening students' agency has been identified as one approach to increasing language learners' learning opportunities (Wagner 2007), for example in ways that enable norms of accountable talk, when students refer to past lessons or their classmates' arguments (Michaels, O'Connor, Hall, and Resnick 2016) or when students explain without being asked (Ingram, Andrews, and Pitt 2019). Schoenfeld's (2013) definition of agency emphasizes active participation in rich activities and students' voices (Wagner 2007).
- *Uses of Assessments*. TRU interprets assessment in informal ways, following the idea that students' reason-

	Mathematical Richness To what extent is the mathematics discussed clear, correct, and well justified (tied to conceptual underpinnings)?	Cognitive Demand To what extent do classroom interactions create and maintain an environment of intellectual challenge?	Equitable Access To what extent do activity structures invite and support active engagement from the diverse range of students?	Agency To what extent do students have opportunities to conjecture, explain, and argue, thus to developing agency and authority?	Use of Contributions To what extent is student reasoning elicited, challenged, and refined?	Discursive Demand To what extent do students engage in rich discourse practices? (additional dimension)	Connecting Registers To what extent are language registers and representations systematically and explicitly connected? (additional dimension)
Level 0	The content is purely rote OR disconnected or unfocused OR consequential mathematical errors or language inaccuracies are not addressed.	Classroom activities are structured so that students mostly apply familiar procedures or memorized facts.	Classroom management is problematic to the point where the lesson is disrupted, OR a significant number of students appear disengaged and there are no overt mechanisms to support engagement.	The teacher initiates conversations. Students' speech turns are short (one sentence or less) and shaped or constrained by what the teacher says or does.	The teacher may note student answers or work, but the student reasoning is not surfaced or pursued. Teacher actions are limited to corrective feedback or encouragement.	No explicit demands to verbalize own ways of thinking, procedures, or solutions OR students only report their processes of calculation.	The content is primarily addressed in only one register/representation OR Different registers/representations are juxtaposed but not related to each other.
Level 1	The content is relatively clear and correct BUT connections between procedures/calculation strategies, concepts, possibly contexts and the meaning-related language are either <i>limited or superficial</i>	Classroom activities offer possibilities of conceptual or language richness or problem-solving challenge, BUT teaching interactions tend to “scaffold away” the challenges and mostly limit students to providing short responses to teacher prompts.	<i>The participation of students is evenly distributed or the teacher gives support so that a variety of students can participate in activities</i> BUT <i>the students do not necessarily carry out higher order activities related to content.</i>	Students have a chance to talk about mathematical content, their own ideas, and meaning-related interpretations BUT “the student proposes, the teacher disposes”: class discussions and student ideas are not explored or built upon.	The teacher refers to student's thinking and student's meaning-related language, perhaps even to common mistakes BUT <i>Ideas with learning potential are not taken as a basis or problematic ideas are not used as challenges.</i>	Students are explicitly asked or are used to explaining meanings and justifying concepts, their own ways of thinking, procedures, and solutions BUT formal and meaning-related language resources are not or incorrectly linked with each other.	Content or tasks are translated into another representation/register BUT changes are always conducted in the same direction.
Level 2	The content is relatively clear and correct AND connections between procedures/strategies, concepts, contexts and meaning-related language are addressed and explained	The teacher's hints or scaffolds encourage and support students in “productive struggle” in building understanding and engaging in mathematical practices or language issues. AND <i>Level of demand is maintained by appropriate scaffolds or prompts.</i>	The teacher actively supports (and to some degree achieves) broad and meaningful participation OR What appear to be established participation structures result in such participation.	Students put forth and defend their ideas and used terminology or meaning-related language. Teacher may ascribe ownership for students' ideas in exposition, OR students respond to and build on each other's ideas.	The teacher solicits student thinking and individual use of meaning-related language AND subsequent instruction responds to those ideas by building on productive beginnings or emerging misunderstanding or language errors.	Students are explicitly asked or are used to explaining meanings and justifying their own ways of thinking, procedures and solutions AND formal and meaning-related language resources are correctly related.	The explicit connection between several registers/representations is stimulated AND realized by verbalizing the connection. OR Changes are conducted flexibly in different directions.

Fig. 2 L-TRU framework: language-responsive mathematics teaching for robust understanding (Adaptations from Schoenfeld's TRU (2013) are marked in grey if they concern language and in *italics* if they were necessary to capture relevant differences in our data set more closely)

ing should be elicited and then worked with formatively. The notion of building upon students' contributions and resources refers to mathematical ideas, but also to language, for instance, by micro-scaffolding (Ingram et al. 2019; Erath et al. 2021).

These considerations underpin the adaptations of the TRU framework into L-TRU in Sect. 3.

3 Adapted and extended framework L-TRU: Language-responsive teaching for robust understanding

As an advance organizer, Fig. 2 presents the L-TRU framework with five adapted dimensions and two new dimensions. For each dimension, three levels of sophistication have been characterized briefly and these are elaborated in Sects. 3.1 and 3.2.

3.1 Adapting the existing dimensions

The following modifications of the five existing dimensions were necessary when adapting the TRU framework into the L-TRU framework for language-responsive mathematics classrooms:

- *Mathematical Richness*. The first dimension is described as “mathematical focus, coherence and accuracy” (in Schoenfeld 2013), or “the mathematics” (Schoenfeld 2018). We chose the term *Mathematical Richness* to characterize the *quality* in the dimension, while still including teachers' mathematical correctness (Schlesinger et al. 2018). With language-responsive instruction in view, the L-TRU adaptation extends the content from rich mathematics (with conceptual focus) to discussing also language aspects such as meanings of concepts.
- *Cognitive Demand*. The second dimension addresses whether students have meaningful opportunities to engage with the content in cognitively challenging ways (Schoenfeld et al. 2018). It includes both teachers' supply and students' use of cognitive challenges (Brühwiler and Blatchford 2011 emphasize the analytic relevance of both supply and use). The demand refers to the mathematical content (whether only procedures are required, or also concepts and their connections) as well as to the degree to which the original demand is maintained or scaffolded away. Maintaining cognitive demand (whether students are engaged in productive thinking processes) has often been identified in qualitative research to co-occur with discursive demand (whether students engage in rich discourse practices such as arguing and explaining) (Herbel-Eisenmann et al. 2011). However, preliminary studies

(Erath and Prediger 2020) suggested that it is worth splitting both dimensions in language-responsive contexts in order to capture subtle differences between collective *thinking* processes and the *discourse practices* in which they engage (see the sixth dimension). In this way, the adapted dimension of *Cognitive Demand* is slightly narrower than Schoenfeld's (2013). To ensure that the dimension can capture relevant differences between observed classrooms, we have augmented the criteria in ‘proficient’ Level 2 to include constantly maintaining the demand (marked in italics in Fig. 2).

- *Equitable Access*. The dimension of equitable access required almost no adaptation in our data with respect to language, but rather with respect to discriminatory power, especially in Level 1. In contrast to TRU, Level 1 is assigned in L-TRU when many students participate even if only with medium demand. This adaptation was necessary better to distinguish the levels in the data.
- *Agency* Schoenfeld's (2013) definition of agency already focused on students' participation in rich practices, so changes were necessary only for including language and substantiating what students talk about in Level 1. Sometimes, a teacher's move can be identified as encouraging agency, but when norms of students' self-initiated interactions are already well established in the classroom culture (Yackel and Cobb 1996), then students can also show agency without explicitly being encouraged.
- *Use of Contributions*. Informal formative assessments and their use are here termed ‘teachers' use of student contributions’, following the idea that students' reasoning and students' language should be elicited and then built upon by challenging prompts and supportive micro-scaffolding (Gibbons 2002; Ingram et al. 2019).

3.2 Two further dimensions

The five adapted quality dimensions of the L-TRU framework cover important aspects of all three domains: the generic quality, the mathematics-specific quality, and (in its adapted version) aspects relevant for promoting language. However, we saw the need to extend the framework by two further dimensions, *Discursive Demand* and *Connecting Registers*:

- *Discursive Demand*. Discourse can refer to many different aspects (Ingram et al. 2019) and is connected to agency. In our conceptualization of discursive demand, we focus on (individually or collectively conducted) *discourse practices*—such as reporting a procedure, explaining the meaning of a concept, and arguing—in which explaining and arguing have been shown to be richer and more difficult for students than reporting procedures (Erath and Prediger 2020; Moschkovich 2015).

However, they are a crucial learning medium for conceptual understanding and should therefore be considered as learning content for language learners (Moschkovich 2015; Pöhler and Prediger 2015). Although qualitative studies have often shown tight connections between rich discourse practices and maintaining cognitive demand and students' agency, we decided to capture separately the *discursive quality* of what students contribute. Only capturing it separately allows for empirical study of the connection. The L-TRU framework therefore captures in a separate quality dimension the richness of discourse practices in which the students engage. An episode is rated Level 0 in *Discursive Demand* when students are not requested to explain their thinking, only report their procedures, or do not explain their thinking in self-initiated ways. Level 1 is assigned when students are requested to explain their thinking or show that they are already used to explaining their thinking or meanings of concepts without (or falsely) connecting the formal and informal language for explaining meanings. A Level 2 rating is given when connections occur. Teachers can maintain high *Discursive Demand*, for instance, by "asking a student to explain someone else's strategy, ... discuss[ing] differences between multiple ideas, ... [or] asking students to connect their own ideas to the ideas of another student" (Ing et al. 2015, p. 348) or by having students explicitly explain or justify concepts based on internalized norms rather than explicit prompts.

- *Connecting Registers*. One instructional approach for developing language learners' conceptual understanding is the use of multiple representations (Zahner et al. 2012) and multiple language registers (Adler and Ronda 2015), which means the connection of everyday language, academic language, and formal language. The degree to which the different representations and registers are not only juxtaposed but deliberately connected is a quality dimension in itself for language learners (Adler and Ronda 2015; Prediger and Wessel 2013). In a comparison between the TRU framework and the MQI framework (Hill et al. 2008), moments where only two representations were juxtaposed were rated more highly in MQI than in TRU, so the TRU framework is already sensitive to the demand for connecting representations rather than juxtaposing them (Schoenfeld et al. 2018). Nevertheless, we still decided to separate this important aspect into its own dimension as it was often emphasized for language learners. In our extended dimension, episodes are rated Level 0 in *Connecting Registers* when only isolated representations or registers stand next to each other, Level 1 when changes between representations or registers are conducted only in one direction, and Level 2 when several registers or representations are *explicitly* related and

used for explaining mathematical or language-related contents.

In these ways, the two extended dimensions correspond to important principles in language-responsive classrooms. The statistical connections to the other dimensions must be determined empirically.

3.3 Required empirical support for the L-TRU framework

Building on existing research, Sects. 2 and 3 reveal *theoretical support* for why the dimensions of the L-TRU framework and its levels might be suitable for capturing relevant differences in language-responsive teaching practices. This support must be substantiated by empirical investigation: The empirical part of the paper uses video data to provide qualitative and quantitative empirical support for the L-TRU framework following three refined research questions:

(RQ1) How can the dimensions capture important differences in teaching practices?

(RQ2) Can we assure sufficient interrater reliability?

(RQ3) How are the quality levels distributed in the episodes of the 41 coded lessons?

4 Methodological framework of the study

4.1 Research context: language-responsive teaching unit for percentages

4.1.1 Design of the teaching unit

The data for the video study stem from the larger project *MuM-Innovation* (*MuM* = *Mathematics learning under conditions of language diversity*), which investigates the implementability of a language-responsive teaching unit aimed at fostering conceptual understanding' of percentages in Grade 7 (Prediger and Neugebauer 2020).

The teaching unit follows the design principles of macro-scaffolding and connecting registers and representations (Pöhler and Prediger 2015):

The *conceptual learning trajectory* of the teaching unit was adapted from Realistic Mathematics Education (van den 2003), which starts with students' everyday experiences and proceeds to constructing meaning for percentages. In Step II, students' informal strategies are elicited and in Step III elaborated into calculation strategies for standard problems. Step IV ends with flexible applications of learned concepts and strategies in more complex contexts (see Fig. 3).

This conceptual learning trajectory is systematically intertwined with a *language-learning trajectory* (Pöhler and Prediger 2015), which starts from students' everyday

	Conceptual learning trajectory towards a mathematical concept (e.g., percentages)	Language learning trajectory for different discourse practices
Step I	Constructing first meanings for the problems and representations (in this study on percentages, in the download bar)	Intuitive use of <i>students' everyday resources</i> for discussing first ideas (e.g., 'how much was downloaded', 'how much is left')
Step II	Developing informal strategies in the context and graphical representation (for determining rates and amounts)	Establish common <i>meaning-related vocabulary</i> for explaining meanings (e.g., 'part of the whole', 'old price', 'new price', 'discount')
Step III	Developing more formal procedures, independent from context & representations (e.g., proportional reasoning)	Introduce <i>formal vocabulary</i> in the technical register for reporting procedures (e.g., 'base', 'amount', 'rate')
Step IV	Applying the mathematical concepts and procedures to more complex problems (also in non-familiar contexts)	Introduce extended reading vocabulary for comprehending texts on non-familiar contexts (e.g., 'VAT', 'net', 'fat content')

Fig. 3 Curricular core of the teaching unit: Intertwined learning trajectories towards percentages

resources by discussing intuitive ideas (Step I), establishes the discourse practice of explaining meanings and supports it by establishing shared meaning-related vocabulary (e.g., old price, new price, rate to be paid) in Step II, using language frames, word banks, and repeated teacher prompts. In Step III, it connects them to formal vocabulary (e.g., quotient, base, amount, rate) for reporting and justifying formal procedures. Step IV finally widens to extended reading demands while cracking more complex percentage problems. The

difference between formal and meaning-related vocabulary has been shown to be crucial, as formal vocabulary is often not sufficient for the discourse practice of explaining meanings. Whereas language proficient students often find their own words for explaining meanings, students with low academic language proficiency have been shown to be in need of sharing meaning-related vocabulary (Pöhler and Prediger 2015).

Task 5. Different offers for favorite shoes

a) Maurice discovered an offer in the city. He draws the offer for his favourite shoes in a percent bar.

Summer Sale
All Sneakers cost only 75 % of the old price!

- What can you see in this bar?
- How has Maurice found the new price of 60 €?
- Compare to Kenan's bar in Task 3.

b) ...

Task 7. Vocabulary for percentage problems

The words on the word cards can help you to describe offers and calculations like in Task 5.

- But which word belongs to what?
- Fill the boxes in the percent bar. Sometimes, more than one word belongs to a box.

Task 13a. How large is the app? (determining the base)

Kenan has already downloaded 2 MB of a comic app. This is 10% of the whole app.

- Include the given information on the percent bar.
- Determine how large the whole app is.
- How many MB must still be downloaded?

Download of "Comic" to "Apps"

Fig. 4 Tasks from the language-responsive curriculum material on percentages

The curriculum material contains 21 instructional tasks for realizing the intended dual learning trajectory and language scaffolds. Three example tasks are shown in Fig. 4 that illustrate how the percent bar mediates the conceptual and lexical aspects as a visual scaffold, following the principle of connecting multiple representations.

4.1.2 Proven effectiveness but differences between classes

In a recent field trial with $n = 655$ seventh graders in 34 classes (Prediger and Neugebauer 2020), the teaching unit showed effectiveness again as the intervention group had significantly higher learning gains than the control group (with $F_{times \times group}(1, 678) = 18, p < 0.001$), even if effect sizes were small ($\eta^2 = 0.036$). The findings suggest that students' conceptual understanding of percentages can be enhanced by the language-responsive teaching unit even for teachers with less proximity to the research group. However, large differences occurred between the intervention classes: The effect of *class adherences* between all filmed groups was measured to be an ICC of 0.27, meaning that 27% of the variance can be explained by the teaching practices in these classes. These differences between classes occurred although all teachers in the intervention classes were prepared by professional development on language-responsive teaching (in four sessions of 3 h each), received curriculum materials in line with a language-responsive instructional approach, and implemented it with a workbook completion rate of more than 75%. Hence, the current video study became necessary to investigate the different teaching practices by which the curriculum was enacted in different classes. In this study, we aimed to identify those dimensions in the teaching practices that differ most.

4.2 Methods of data gathering of the video study

The video data corpus of the study comprises 2,520 min of video from 41 lessons held by 26 mathematics teachers in their Grade 7 classes (age 12). The percentage of multilingual students in these classes ranged from 30 to 90% and academic language proficiency was mostly low to medium (Prediger and Neugebauer 2020). The teachers' teaching experience span was from 5 to 20 years. All 26 teachers held a teaching certificate in secondary mathematics, but received different amounts of professional development on language support in mathematics classrooms. To ensure comparability, all received the language-responsive curriculum material for percentages (presented in Sect. 4.1).

4.3 Methods of data analysis of the video study

To rate the data, we followed Schoenfeld's (2018, p. 496) procedure to segment each video-taped lesson into episodes

of up to 5 min of the same activity type (whole-class discussion, student presentation, small group work, and individual work). A new episode was also segmented with a new task.

Each of the 497 segmented episodes was rated on a three-point scale (Level 0 = *basic*, 1 = *proficient*, 2 = *distinguished*). To ensure the reliability of the rating process, a rating protocol with flow charts was iteratively refined in the initial coders' collective discussions. Two independent coders then rated all material to determine interrater reliability.

To visualize the course of levels within a lesson, lesson profiles were printed on timelines for each dimension. The lesson profiles reveal a quick overview of the course of the ratings in relation to treated tasks and activity structures.

Since the rating was conducted on an ordinal scale, all methods of analysis must correspond to the ordinal level of measurement. The descriptive analysis of the data summarized the frequencies of ratings on each level and compared them for all teachers and for teachers with the highest and lowest overall ratings. Statistical connections between the dimensions were determined using Spearman's ρ rank coefficient, which measures association between ordinal scales for small sample sizes. Non-parametric ρ tests were administered to test the statistical dependences of the scales.

5 Qualitative empirical insights into the variance of teaching practices with the same language-responsive curriculum material

In this section we pursue research question RQ1 (*How can the dimensions capture important differences in teaching practices?*) in a qualitative way. Section 5.1 shows two episodes on the same task that differ in nearly all dimensions in their rating. Section 5.2 shows relevant differences between the dimensions *Cognitive Demand* and *Discursive Demand*. Zooming out, Sect. 5.3 shows how the three episodes are embedded in the complete lesson profiles, illustrating typical variances within and between lessons.

5.1 Same task, different teaching practices

In Episodes 1 and 2, the classes of the teachers with pseudonyms Mandy and Peter worked on Task 7 (from Fig. 4), asking the students to systematize previously introduced meaning-related vocabulary ('discount', 'share', and 'whole') by assigning them to the percent bar.

The transcript from Episode 1 shows how Mandy clarified the meaning of 'discount' before her students started individual work on Task 7.

Transcript from Episode 1: Mandy's clarification of the concept 'discount'

- 1 Teacher So, we start with Task 7. In there is a word that you all knew last Friday; [...] which important word is that?
- 2 Pia Discount.
- 3 Teacher Yes, discount, isn't it? [...] Who can repeat again, what is behind that word? [5-second break] Nigel?
- 4 Nigel Um, reduced from the price, though, cheaper.
- 5 Teacher OK, the price gets cheaper, when there is discount, doesn't it? And we had a nice example from our everyday life [...] Which example was that? [to Pia].
- 6 Pia Um, for example, winter sale.
- 7 Teacher Winter sale, yes, but that is not what I meant, now. One that fits into our class. Lily.
- 8 Lily Calculator.
- 9 Teacher Lily had bought a calculator [but had not bought it as a part of the centralized order of the school].... So Lily had to pay the full price in the shop. ... [Whereas we have] received a quantity discount of seven euros. A lot of money.

The transcript from Episode 2 shows how Peter's class negotiated the meaning of 'discount' in a whole-class discussion after individual work on the task.

Transcript from Episode 2: Peter's collective negotiation of meaning

- 1 Teacher OK, then we have to discuss it [where to put the cards]. Let us start with discount, that was first. Some thought it belongs at 75% [teacher hints at the 75% on the percent bar] and others thought it belong here in between [hints to the bar between 75 and 100%].
- [5-seconds break] Are there proponents for this 75%? It came from whom?
- 2 Elena From us?
- 3 Teacher What was your or your group's idea?
- 4 Lena Well, Elena thinks that 25, um, 75% calculated, um [5-seconds break] well, these, on this, it was reduced, 25% made.
- 5 Teacher Uh-huh [nodding, 5-seconds break] Would someone like to support that or would you be opponents?
- 6 Students No, no.
- 8 Milan Um, there is discount, though, because, um, it means, um, offer, that, um, 25% are on, um,

- away from 100%. Thus, that is, you do not pay 75%.
- 9 Teacher Dorian.
- 10 Dorian I say it is right because, um, that is the part that you save thus it is, um, part and discount, I see them as equal.
- 11 Teacher Uh-huh [nodding].
- ...
- 14 Tabea Well, um, you shall pay 75% and not save, so this here is correct, on the left.
- 15 Teacher OK. Are you convinced, Lara, or can you object again?
- 16 Lena Am convinced.
- 17 Teacher What has convinced you, actually?
- 18 Lena Yes, um, that you have, um, that you have to pay 75% and that is, though, that not reduced.
- 19 Teacher Yeah, the number was said already. How many percent do you save? You said that in the beginning.
- 20 Lena 25%.
- 21 Teacher Exactly, and the 25% [hints at the percent bar places for discount and 75%] or is that the discount or what you have to pay?

Although both episodes involving Task 7 have the same aim (clarifying the meaning of 'discount'), the comparison shows how different the teaching practices can be:

- In both transcripts, terms are clarified by connecting academic language to everyday language, hence the dimension *Connecting Registers* is equally rated as Level 1.
- However, the *Discursive Demand* has large differences, as students provide only single words in Episode 1 (Level 0) and rich argumentations in Episode 2 (Level 2).
- In the transcribed excerpt of Episode 1, students serve only as keyword providers whose contributions are not really followed up (*Use of Contributions* Level 0 in the transcript, but Level 1 during the whole 5-min episode), whereas in Episode 2, the students' contributions are encouraged, discussed, and built upon (Level 2).
- This is also reflected in the dimension *Agency* for which there is no space in Mandy's classroom (Level 0) but much encouragement in Peter's classroom (Level 2) as students work with others' ideas.
- The difference is smaller in the dimension *Equitable Access*: Whereas Mandy includes some students, but not in rich activities (Level 1), Peter provides access for many different students (Level 2).
- Although the content is rich, in principle (*Mathematical Richness* Level 2 in both transcripts), Mandy does not maintain a high *Cognitive Demand* (Level 0), as students do not start thinking for themselves, whereas Peter

engages the students in productive struggles about the relation of discount and new price (Level 2).

156 Laura Laura OK.
 157 Teacher Teacher That is a sufficient explanation, for the moment.

5.2 Difference between cognitive demand and discursive demand

Many qualitative studies suggest that *Cognitive* and *Discursive Demand* coincide (see Sect. 3.1), and this is also theoretically plausible because the discourse practices are a medium for communicating about cognitive challenges. This is also true for many of our rated episodes.

However, the following example shows that there can be a substantial theoretical and practical difference. In the transcript from Episode 3, Ricardo works with his class on Task 13a (from Fig. 4), asking students for the first time to determine the base (If 10% of the app is 2 MB, how many MB is the whole app?). This episode is mathematically demanding, and Ricardo starts by eliciting explanations. He then maintains the cognitive demand while reducing the discursive demand.

Transcript from Episode 3: Ricardo’s support for structuring the percent bar

144 Laura Here is 0%, here is 100% [hints at the percent bar]. If you count here, 18 MB.
 145 Teacher Uh-huh [nodding].
 146 Laura Laura The 18, do we have to times or divide?
 147 Teacher First, first the labels are right. What would you do now?
 148 Laura Now I have to know this, how many MB?
 149 Teacher Uh-huh [nodding].
 150 Laura Yes, it must be 2.
 151 Teacher Uh-huh [nodding]. Why must it be 2?
 152 Laura Because, the whole is 20.
 153 Teacher How do you know?
 154 Laura LauraBy them [gestures on her work sheet that the 2 MB fit 10 times into the whole].
 155 Teacher Teacher By them, OK, well done.

The rating of the complete Episode 3 is also listed in Table 1. In the overall 5-min period,

- the teacher involves many students and provides access for underprivileged students such as Laura (*Equitable Access* Level 2),
- with medium *Agency* (Level 1),
- he makes distinguished *Use of Students’ Contributions* (Level 2) by referring to Laura’s thinking (Turn 147),
- while *Connecting Registers* (Level 1) in a proficient but not distinguished (Turn 144) way.

It is specifically interesting to see how he maintains high *Cognitive Demand* (Level 2) while discussing how to measure the whole bar with the colored 10%. Laura cannot explicitly explain her idea that the 2 MB fits 10 times into the whole bar, but she can show it by an embodied action (Turn 154). So she dives into deep mathematics without articulating her explanation beyond the embodied action (*Discursive Demand* Level 1), which completely clarifies the mathematical structure in view, hence maintains the

Table 1 Ratings in seven dimensions for the 5-min episodes from which the transcripts were taken

	Episode 1 (Mandy)	Episode 2 (Peter)	Episode 3 (Ricardo)
Mathematical richness	2	2	1
Cognitive demand	0	2	2
Equitable access	1	2	2
Agency	0	2	1
Use of contributions	1 (0 in the transcript)	2	2
Discursive demand	0	2	1
Connecting registers	1 (0 in the transcript)	1	1

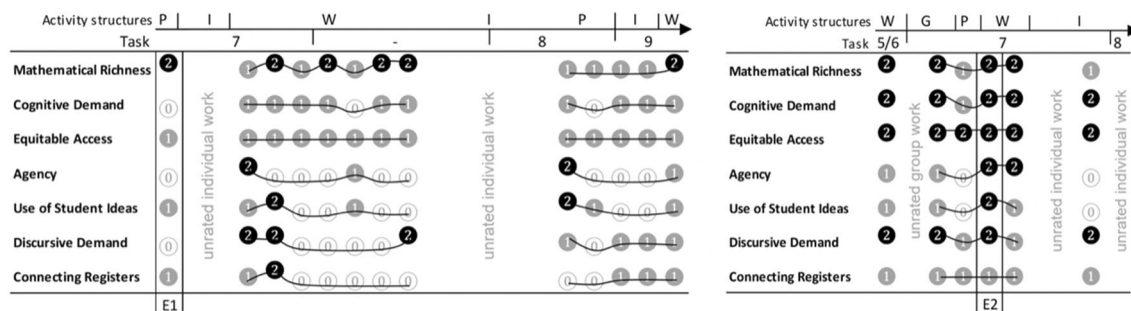


Fig. 5 Lesson profiles for Mandy’s (left) and Peter’s (right) lesson (Episodes E1 and E2 are marked by rectangles; W whole class, G group work, P student presentation, I individual work)

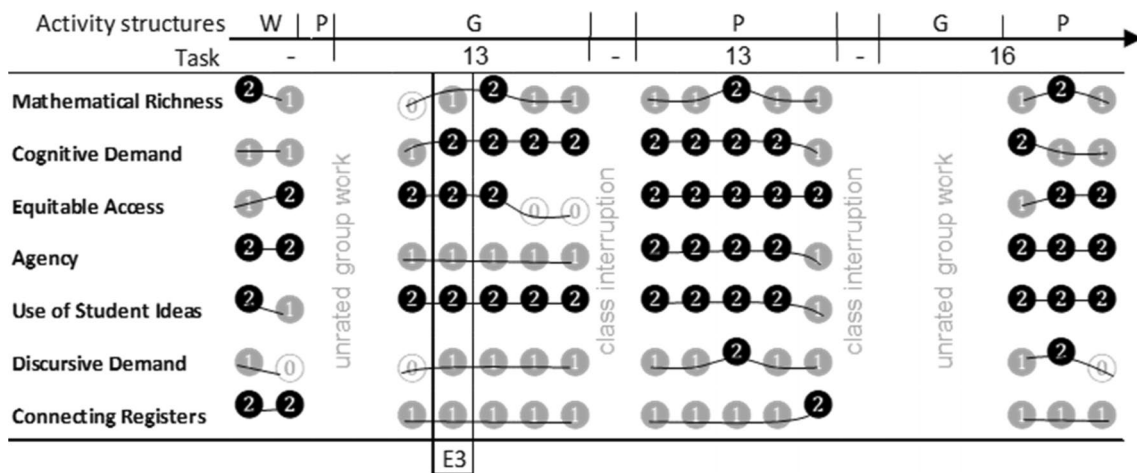


Fig. 6 Lesson profile for Ricardo’s lesson (Episode E3 is marked by a rectangle; W whole class, G group work, P student presentation, I individual work)

Cognitive Demand Level 2 that Ricardo values (T 153) by asking already answered questions (Ingram et al. 2019) on a deeper level (Turns 147, 151 & 153). This moment of disentangling *Cognitive* and *Discursive Demand* is possible due to Ricardo’s acceptance of gesturing and support for articulating.

The ratings of the 5-min episodes from which these transcripts provide excerpts are summarized in Table 1. Although two of the ratings of the transcripts differ from the overall ratings of the 5-min segments, overall validity has been shown to be sufficient with segments consisting of 5-min episodes (Schoenfeld 2018).

5.3 Embedding the three episodes in the corresponding lesson profiles

The comparison of the transcripts to the complete 5-min episodes has already revealed that the quality dimensions were not constant. The variation within lessons is visible in Figs. 5 and 6, where the three episodes are located in profiles of the whole lessons (embedding visualized using rectangles E1, E2, and E3). In this way, the lesson profiles provide a good overview of the varying quality of the enacted teaching practices.

The comparison shows that lessons always cover different levels for most of the seven dimensions across the different activity structures. Only *Equitable Access* seems to be more constant. These observations call for a quantitative analysis of all ratings, on which we report in Sect. 6.

6 Quantitative descriptions of the quality dimensions

6.1 Interrater reliability

In order to pursue quantitative analyses of instructional practices using video material, reliability must be ensured. As the interrater reliability has not yet been analyzed for the TRU framework (Schoenfeld 2018, p. 496), it was investigated in research question RQ2.

To determine the interrater reliability, 1610 min (2/3 of the total video material) was rated by two independent raters. The ratings never deviated more than one level. Table 2 shows that Cohen’s κ is sufficient or good in all dimensions, with an overall $\kappa=0.78$. After determining the interrater reliability, all moments of disagreement were resolved by consensus so that the successive analyses could be based on 100% rater agreement.

Table 2 Interrater reliabilities in each dimension

	Cohen’s κ
Mathematical richness	0.86
Cognitive demand	0.81
Equitable access	0.88
Agency	0.73
Use of Contributions	0.69
Discursive demand	0.75
Connecting registers	0.79
Over all agreement	0.78

Table 3 Association between the quality dimensions: Spearman's rank correlation for all 5-min episodes

Spearman's ρ	Mathematical richness	Cognitive demand	Equitable Access	Agency	Use of Contributions	Discursive Demand	Connecting Registers
Mathematical richness	1	0.39	0.28	0.29	0.35	0.50	0.22
Cognitive demand		1	0.37	0.40	0.31	0.40	0.26
Equitable access			1	0.30	0.18	0.33	0.12
Agency				1	0.52	0.44	0.33
Use of contributions					1	0.40	0.24
Discursive demand						1	0.25
Connecting registers							1

6.2 Dependencies between quality dimensions

A further step of analyzing the usability of the L-TRU framework re-addressed research question RQ1 (*How can the dimensions capture important differences in teaching practices?*) not only qualitatively (see Sect. 5) but also quantitatively by analyzing the correlations between dimensions. Schoenfeld (2018, p. 493) described his five dimensions as theoretically independent, but some directed dependencies are theoretically plausible, for instance, no *Cognitive Demand* without *Mathematical Richness*. There are also case studies showing how *Cognitive and Discursive Demand* coincide, or *Agency and Discursive Demand* (see Sect. 3.2). These qualitative findings can now be triangulated by quantitatively investigating the association.

Table 3 reveals the Spearman's rank correlation for all 497 episodes of 5 min each. Due to the limited range of 0 to 2, all rank correlations are significant (most pairs with $p < 0.01$, only *Equitable Access* correlates with *Use of Contribution* and *Connecting Registers* with $p < 0.05$).

However, as the correlations range between 0.12 and 0.52, most dimensions can be considered independent, so they can indeed capture different aspects of the teaching practices. Especially the new dimension *Connecting*

Registers correlates with other dimensions only with ρ between 0.12 and 0.33. Stronger associations exist between other dimensions:

- The association between *Mathematical Richness* and *Cognitive Demand* ($\rho = 0.39$) confirms the fact that *Cognitive Demand* requires *Mathematical Richness*, but they are still not identical, for instance, if a demand is too high so that students cannot engage.
- The association between *Cognitive* and *Discursive Demand* (0.40) was expected from the literature because cognitively demanding discussions usually require elaborated discourse practices (see Sect. 3). However, the association is lower than for three other dimensions, and Episode 3 gave an example of how a teacher can give relief from high *Discursive Demand* by gesturing or micro-scaffolding.
- *Mathematical Richness* has the second strongest association to *Discursive Demand* (0.50), which is more in line with prior theoretical assumptions.
- The association between *Equitable Access* and *Cognitive Demand* (0.37) can be traced back to overlapping operationalizations of their Level 2 ratings: Level 2 of *Equitable Access* calls not only for wide participation but participation in cognitively demanding activities (see

Table 4 Distribution of quality levels for all teachers and weakly and strongly rated teachers

Episodes rated as in dimension	...of all teachers			... of 3 teachers with highest ratings			... of 3 teachers with lowest ratings		
	Level 0	Level 1	Level 2	Level 0	Level 1	Level 2	Level 0	Level 1	Level 2
Mathematical Richness	0%	49%	51%	0%	22%	78%	0%	56%	44%
Cognitive demand	3%	25%	71%	0%	11%	89%	9%	47%	44%
Equitable access	1%	15%	84%	3%	6%	91%	3%	50%	47%
Agency	14%	40%	46%	7%	23%	71%	41%	32%	27%
Use of contributions	24%	39%	37%	11%	22%	67%	41%	44%	15%
Discursive demand	25%	45%	30%	3%	37%	60%	35%	32%	32%
Connecting registers	9%	70%	21%	2%	61%	37%	44%	47%	9%

Distribution of episodes were rated on each level in seven dimensions. Sums below 100% occurred when episodes were rated as 'not applicable'

Fig. 2). However, both dimensions are important because (after the adaptation in L-TRU), Level 0 and Level 1 concern classroom management in the dimension *Equitable Access*. So distinguished levels require each other, but the lower levels are more independent. The same applies for *Cognitive Demand* and *Agency* (0.40).

- Also the strong association between *Agency* and *Use of Contributions* (the strongest association, at 0.52) can be explained theoretically as *Agency* Level 2 being operationalized by making use of students' contributions. However, both dimensions do not capture the same thing since a distinguished *Use of Contributions* is one pathway towards distinguished *Agency*, but *Agency* Level 2 can also be reached by *Use of Contribution* Level 1, especially in group work settings.

6.3 Descriptive data on distribution of quality levels

To pursue research question RQ3 (*How are the quality levels distributed in the episodes of the 41 coded lessons?*), the quality-level ratings of 497 episodes were counted. Table 4 shows the distribution of quality levels for all teachers in the first three columns. The other columns provide the frequencies for the three teachers with the lowest ratings and those for the three teachers with the highest ratings. Mandy from Episode 1 belongs to the teachers with the lowest ratings, Peter from Episode 2 to those with the highest ratings and Ricardo from Episode 3 to the medium group.

For four dimensions, all observed teachers maintained Level 1 or above. In particular *Mathematical Richness* was only rated Level 0 in 4 out of 497 episodes (0%), which means that in 493 episodes mathematical or language-related mistakes were addressed adequately. Also *Equitable Access* Level 0 occurs only in 1% of the episodes, which means that the observed teachers had productive practices for classroom management.

The three dimensions with the highest percentage of Level 0 are *Use of Contributions* (24%), *Discursive Demand* (25%), and *Agency* (14%). These dimensions seem to be more challenging, presumably due to the required adaptivity in teachers' facilitation.

In contrast, *Cognitive Demand* (71%) and *Equitable Access* (84%) have the highest percentage in Level 2. These dimensions seem to be well mastered by many teachers.

Connecting Registers is the only dimension with more Level 1 (70%) than Level 2 (21%) ratings. This means that changes of registers and representations are often initiated but not necessarily permanently connected explicitly or flexibly, as was expected.

When considering only average distributions for *all* teachers, the teaching practices cannot be traced back to the

curriculum material. The comparison of teachers with the lowest and highest ratings (in the six columns on the right of Table 4) provides a first access to identifying what the curriculum material can support best: When a dimension is rated highly for teachers with the lowest overall ratings, the curriculum material might contribute to achieving this quality more than for dimensions with larger differences.

The teachers with high overall ratings seem to outperform the teachers with low overall ratings, achieving Level 2 ratings particularly in the dimensions *Cognitive Demand* (89% vs. 44%, respectively), *Discursive Demand* (60% vs. 32%, respectively), and *Use of Contributions* (67% vs. 15%, respectively). These dimensions seem to have the most dependence on the teachers' individual enactment, with the highest distance for *Discursive Demand*, the quality dimension which is most emphasized as relevant for language learners (Herbel-Eisenmann et al. 2011; Moschkovich 2015). On the other side of the scale, 10 times fewer teachers with high ratings were assigned Level 0 (3%) for *Discursive Demand* than teachers with low ratings (35%) were. This shows that the new dimension *Discursive Demand* seems to capture substantial differences. The other new dimension, *Connecting Registers*, also reveals large differences between teachers with high and low ratings: 2% vs. 44% on Level 0 and 37% vs. 9% on Level 2, respectively. Both newly introduced dimensions therefore seem to capture visible differences that were not addressed by the previous dimensions.

The comparison of the distribution between teachers with high and low ratings for *Equitable Access* (6% vs. 50% on Level 1 and 91% vs. 47% on Level 2, respectively) shows that all teachers were able to involve many students (Level 1), but the teachers with higher overall ratings involved the students more often in demanding activities (Level 2).

7 Discussion

7.1 Summary

As Erath et al. (2021) described in the introduction of this special issue, design principles for language-responsive learning environments and teaching practices are often treated separately in the literature. However, qualitative studies have indicated that the teachers' enactment of well-designed curriculum material can massively influence the quality of the students' learning opportunities (Gibbons 2002; Adler and Ronda 2015) and that students with limited access to academic language might be the most vulnerable for low-quality learning opportunities (Herbel-Eisenmann et al. 2011). Although qualitative studies have often shown these opportunity gaps, no previous framework has combined insights about language learners' learning needs with the research on generic quality instruction and

mathematics-specific quality instruction (Charalambous and Praetorius 2018; Erath and Prediger 2020).

In this paper, we are trying to reduce this research gap by developing a framework for quantitatively assessing the quality of teaching practices. For this purpose, we adapted and extended Schoenfeld's (2013, 2018) TRU framework into the L-TRU framework for language-responsive mathematics teaching for robust understanding. To show that the L-TRU framework is a suitable research tool, we argued *theoretically* how the dimensions build upon former research. Many theoretical assumptions could also be made for possible associations of dimensions that are surely not completely independent. However, the associations were also studied empirically rather than anticipated theoretically.

The *empirical substantiation* drew upon video data from 41 language-responsive Grade 7 lessons by teachers who worked with the same language-responsive curriculum material (Pöhler and Prediger 2015). The first analyses in the ongoing project have allowed us to accomplish the following:

- (RQ1) show that the dimensions capture important differences (qualitatively in transcripts and by correlations between dimensions), even in classrooms with the same curriculum materials; project have allowed us to
- (RQ2) report measures for sufficient and good interrater reliability ($\kappa=0.78$); and.
- (RQ3) provide descriptive data on the distribution of quality levels in different dimensions.

The five dimensions from Schoenfeld's TRU framework (2013) required only minimal adaptations for language-responsive classrooms as they already covered important aspects (distinct from those of the new scales).

Special attention was dedicated to the two additional dimensions *Discursive Demand* and *Connecting Registers*. *Connecting Registers* was the dimension with the lowest rank correlation to other dimensions (between 0.1 and 0.26). As its importance for language learners has often been emphasized (Zahner et al. 2012; Prediger and Wessel 2013), it might therefore be relevant to capture it separately. *Discursive Demand* has often been described as tightly connected to *Cognitive Demand* and *Mathematical Richness* in qualitative studies (Herbel-Eisenmann et al. 2011; Erath and Prediger 2020). Its separation in a new dimension allowed us to confirm this hypothesized association statistically ($\rho=0.40$ and $\rho=0.50$, respectively), but *Agency* and *Use of Contributions* have a higher association (0.52). Additionally, *Discursive Demand* turned out to be the most discriminating dimension, meaning that it is the one that shows the greatest difference between teachers with low overall ratings and those with high overall ratings. The fact that all associations

were below 0.52 makes it clear that care should be taken with statements of equivalence that are too simple.

On the other hand, the exact values of the rank correlations should not be over interpreted, as they might be strongly tied to this particular teaching unit with its prepared curriculum material; therefore, they may not yet be generalizable.

Overall, the first analysis of 41 lessons shows that while the curriculum material seems to assure a medium quality in the dimension *Connecting Register* and a high quality in the dimensions *Cognitive Demand* and *Equitable Access*, the dimensions *Agency*, *Discursive Demand* and *Use of Contributions* show the largest variance. Among all dimensions, *Discursive Demand* seems to show the most separation between teachers with high and low overall ratings.

7.2 Limitations and future research

Although empirical support for the development of the L-TRU framework could be provided, the general nature of the results should not be overestimated, as they rely only on 41 lessons held by 26 teachers using the same curriculum material in one teaching unit. Future research will have to investigate whether the data set was representative and whether these results also hold for classes working with different curriculum materials.

More importantly, the current data analysis could show only that the L-TRU framework can capture differences between teaching practices in seven dimensions. As long as we have not identified which of these dimensions really correlate with students' learning gains, the claim of having identified *quality* dimensions is empirically supported only by the cited qualitative studies, not yet by quantitative investigations of impact of the dimensions on effectiveness (Charalambous and Praetorius 2018; Brophy 2000). Hence, the next step of the project will relate the L-TRU ratings to students' learning gains and investigate their impact. After this step, the plan is to use the analytic framework in research in the supply-use model, which also involves students' aptitudes (Brühwiler and Blatchford 2011).

7.3 Practical consequences for implementing language-responsive instructional approaches

Although the proof of effectiveness is still pending, the state of qualitative research combined with our new findings on the frequency of occurrences can already suggest practical consequences for implementation projects for language-responsive instructional approaches. Providing well-designed curriculum material is an effective implementation strategy, but the curriculum material alone can

ensure only some quality dimensions, whereas others need teachers' expertise in enacting the approaches. These findings suggest that professional development should sensitize teachers to the following:

- their already existing expertise in the dimensions *Equitable Access* and *Cognitive Demand*, which can now be extended to language-learning content;
- huge variances between teaching practices with respect to *Agency*, *Discursive Demand*, and *Use of Contributions* (and ways to enhance these dimensions, as addressed by Michaels et al. 2016);
- the high relevance of *Discursive Demand* (Herbel-Eisenmann et al. 2011; Moschkovich 2015); and
- the difference between juxtaposing multiple representations and registers and really explicitly *Connecting Registers* (Adler and Ronda 2015; Prediger and Wessel 2013).

These aspects can be realized, for instance, by discussing video clips and their ratings in professional development (Schoenfeld 2018). Additionally, the curriculum material can be enriched with more prompts that have high *Discursive Demand* in order to give even greater support to teachers.

Acknowledgements The study was conducted in the *MuM-Innovation* project, supported by the National Ministry of Education and Research under BMBF-Grant 03VP02270 to S. Prediger, in the research framework of DZLM, the German Center for Mathematics Teacher Education (funded by German Telekom Foundation).

Funding Open Access funding provided by Projekt DEAL.

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

References

- Adler, J., & Ronda, E. (2015). A framework for describing mathematics discourse in instruction and interpreting differences in teaching. *African Journal of Research in Mathematics, Science and Technology Education*, 19(3), 237–254.
- Brophy, J. (2000). *Teaching (Educational Practices Series Vol. 1)*. Brussels: International Academy of Education (IAE).
- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21(1), 95–108.
- Brunner, E. (2018). Qualität von Mathematikunterricht: Eine Frage der Perspektive. *Journal für Mathematik-Didaktik*, 39(2), 257–284. <https://doi.org/10.1007/s13138-017-0122-z>.
- Callahan, R. M. (2005). tracking and high school english learners: limiting opportunity to learn. *American Educational Research Journal*, 42(2), 305–328.
- Charalambous, C. Y., & Praetorius, A.-K. (2018). Studying mathematics instruction through different lenses. *ZDM*, 50(3), 355–366.
- DIME—Diversity in Mathematics Education Center for Learning and Teaching. (2007). Culture, race, power in mathematics education. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 405–433). Charlotte: Information Age.
- Erath, K., & Prediger, S. (2020). Quality dimensions for activation and participation in language-responsive mathematics classrooms. In N. Planas, M. Schütte, & C. Morgan (Eds.), *Classroom research on mathematics and language - seeing learners and teachers differently*. London: Routledge. (in press).
- Erath, K., Ingram, J., Moschkovich, J., & Prediger, S. (2021). Designing and enacting teaching that enhances language in mathematics classrooms. *ZDM Mathematics Education*, 53(2), in this issue.
- Gibbons, P. (2002). *Scaffolding language, scaffolding learning. Teaching second language learners in the mainstream classroom*. Portsmouth: Heinemann.
- Henningesen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition. *Journal for Research in Mathematics Education*, 28(5), 524–549.
- Herbel-Eisenmann, B., Choppin, J., Wagner, D., & Pimm, D. (2011). *Equity in discourse for mathematics education*. Dordrecht: Springer.
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte: Information Age.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., et al. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction. *Cognition and Instruction*, 26(4), 430–511.
- Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., et al. (2015). Student participation in elementary mathematics classrooms. *Educational Studies in Mathematics*, 90(3), 341–356.
- Ingram, J., Andrews, N., & Pitt, A. (2019). When students offer explanations without the teacher explicitly asking them to. *Educational Studies in Mathematics*, 101(1), 51–66.
- Jackson, K., Gibbons, L., & Sharpe, C. (2017). Teachers' views of students' mathematical capabilities. *Teachers College Record*, 119(7), 1–43.
- Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. B. (2016). *Accountable talk sourcebook: For classroom conversation that works*. Pittsburgh: Institute for Learning. <https://ifl.pitt.edu/how-we-work/sourcebook.cshml>.
- Moschkovich, J. (2015). Academic literacy in mathematics for english learners. *Journal of Mathematical Behavior*, 40(A), 43–62.
- Pöhler, B., & Prediger, S. (2015). Intertwining lexical and conceptual learning. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(6), 1697–1722. <https://doi.org/10.12973/eurasia.2015.1497a>.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of three basic dimensions. *ZDM*, 50(3), 407–426.
- Prediger, S., & Neugebauer, P. (2020, submitted). Can students with different language backgrounds equally profit from a language-responsive instructional approach for percentages? *Submitted Manuscript*.

- Prediger, S., & Wessel, L. (2013). Fostering German language learners' constructions of meanings for fractions. *Mathematics Education Research Journal*, 25(3), 435–456.
- Schlesinger, L., Jentsch, A., Kaiser, G., König, J., & Blömeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM*, 50(3), 475–490.
- Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM*, 45(4), 607–621.
- Schoenfeld, A. H. (2018). Video analyses for research and professional development: the teaching for robust understanding (TRU) framework. *ZDM*, 50(3), 491–506.
- Schoenfeld, A. H., Floden, R., Chidiac, F. E., Gillingham, D., Fink, H., Hu, S., et al. (2018). On classroom observations. *Journal for STEM Education Research*, 1(1–2), 34–59.
- Secada, W. G. (1992). Race, ethnicity, social class, language and achievement in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 623–660). New York: MacMillan.
- van den Heuvel-Panhuizen, M. (2003). The didactical use of models in realistic mathematics education: an example from a longitudinal trajectory on percentage. *Educational Studies in Mathematics*, 54(1), 9–35.
- Wagner, D. (2007). Students' critical awareness of voice and agency in mathematics classroom discourse. *Mathematical Thinking and Learning*, 9(1), 31–50.
- Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, 27(4), 458–477. <https://doi.org/10.2307/749877>.
- Zahner, W., Velazquez, G., Moschkovich, J., Vahey, P., & Lara-Meloy, T. (2012). Mathematics teaching practices with technology that support conceptual understanding for Latino/a students. *Journal of Mathematical Behavior*, 31(4), 431–446.

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