



# Types of mathematical tasks in lower secondary classrooms in Germany

## Statistical findings from a latent class analysis based on general mathematical competencies

Ann-Kristin Adleff<sup>1</sup> · Natalie Ross<sup>1</sup> · Johannes König<sup>2</sup>  · Gabriele Kaiser<sup>1,3</sup> 

Accepted: 27 June 2023 / Published online: 11 September 2023  
© The Author(s) 2023

### Abstract

Tasks play an important role in mathematics education, as they provide opportunities for students to develop their competencies and to cognitively engage with the mathematical content. The potential for cognitive activation as a central feature of a mathematical task has been considered in numerous studies, mostly as a didactical analysis by means of existing classification systems. In this study, which has been carried out in the frame of the TEDS research program, a different approach is taken by which we aim to identify the cognitive demand level of tasks used in ordinary mathematics lessons. Thus, the correlation between general mathematical competencies and the cognitive mathematical activities required to solve the task was analyzed and common properties of groupings of tasks were empirically extracted. In detail, 2490 tasks from mathematics lessons in Germany were analyzed by means of a rational task analysis with regard to their potential for fostering general mathematical competencies, namely modelling, problem solving, reasoning and argumentation, use of representations, use of symbols and operations, and communication. Latent class analysis revealed six classes of mathematical tasks with varying potential for the different competencies. In accordance with previous studies on mathematical tasks in Germany, the biggest class focused solely on the use of symbols and operations, while other classes showed different foci. Post hoc analyses revealed that the classes of tasks differ with regard to the level of cognitive activity they require. The results of the study highlight that the potential for cognitive activation of the tasks used in the classrooms of this more recent study has not improved in the last decades, despite many reform activities in German mathematics education, and that many mathematical tasks used are still more or less calculation oriented.

**Keywords** Mathematical tasks · Latent class analysis · Mathematical competencies · Rational task analysis · Potential for cognitive activation

## 1 Introduction

The selection of mathematical tasks for students to work on during a lesson is a powerful tool for shaping mathematics education; in fact, it is “among the most important decisions a teacher makes” (Boston & Smith, 2011, p. 966). The planning of learning tasks integrates numerous didactic decisions made by the teacher, including those regarding the goals, content, and activities of the lesson (König et al., 2020). Watson and Ohtani (2015) highlight the importance of tasks for mathematics education:

From a cognitive perspective, the detail and content of tasks have a significant effect on learning; from a cultural perspective, tasks shape the learners’ experience of the subject and their understanding of the nature of mathematical activity; from a practical perspective, tasks are the bedrock of classroom life, the “things to do.” (p. 3)

Tasks not only determine the mathematical content and processes with which students interact during the lesson but also impact the situation in which content is embedded (Doyle, 1983). In an extension of the well-known didactical triangle to a socio-didactical tetrahedron, Rezat and Sträßer (2012) consider lesson artifacts<sup>1</sup> as a fourth and fundamental constituent of the didactical situation. In each of the new triangles (student–task–teacher, student–task–content, and teacher–task–content), tasks shape the relations between the other two corner points. For instance, tasks provide students with the opportunity to actively engage with mathematical content while enabling teachers to structure said content.

On an intuitive level, it seems plausible that high-quality tasks facilitate (and indicate) high-quality teaching (as defined by Berliner, 2005), which in turn leads to better student learning. Many studies have investigated the effects of high-quality mathematics instruction on cognitive or motivational student outcomes (e.g., Baumert et al., 2010; Blömeke et al., 2022; Hiebert & Grouws, 2007; Lynch et al., 2017; Scherer et al., 2016) and have mostly found small to moderate positive effects. Other studies have discussed the possible impact of high-quality tasks on students’ cognitive learning gains and achievement (English, 2011; Sullivan & Mornane, 2014), as well as students’ motivation (Heinle et al., 2022), but strong empirical evidence is still missing. Still other studies, particularly those conducted on classroom artifacts, point to a significant correlation between task characteristics and the quality of instruction (Hill & Charalambous, 2012; Matsumura et al., 2002, 2006). As artifacts are relatively easy to sample and can be re-analyzed at a later stage if needed, the potential of tasks is often used as a direct indicator of the quality of instruction (Baumert et al., 2010; Boston, 2012; Herbert & Schweig, 2021). However, few studies have explored in detail the assumed relations between different facets of task potential and instructional quality (e.g., Herbert & Schweig, 2021; Joyce et al., 2018). Recent studies have adopted interesting approaches that relate task characteristics to students’ mathematical creativity (Levensen et al., 2018; Lithner, 2017) or to global goals like the promotion of peace and sustainability (Yaro et al., 2020). Such work highlights the importance of considering the task perspective for various goals of mathematics education.

Despite the high relevance of mathematical tasks to both research and practice, the field lacks a comprehensive understanding of the kinds of tasks that are used in ordinary classrooms and how they impact the teaching and learning of mathematics. Therefore, the aim of

<sup>1</sup> In the analysis, the terms *lesson artifact* and *classroom artifact* refer not only to the tasks used throughout the course of a lesson but also include other objects, such as students’ notes or lesson plans.

this paper is to empirically identify and characterize different types of tasks used in common classrooms based on an in-depth analysis of task features relevant for mathematical learning processes. This approach is in contrast to other studies, in which the distribution of the characteristics of the tasks within a task sample is analyzed with an existing classification system.

## 2 Tasks in mathematical educational research and practice

According to the widely accepted work by Doyle (1983), academic tasks “are defined by the answers students are required to produce and the routes that can be used to obtain these answers” (p. 161). These routes are determined by the operations students are required to perform as well as the resources they have at their disposal. Johnson et al. (2017) view the task as “a form of social practice, undertaken by teacher and students as a collective” (p. 814). Both perspectives on tasks can be integrated with the mathematical task framework proposed by Stein et al. (1996), which illustrates three phases of task implementation. First, the mathematical task is considered as it is presented in the material. Second, the teachers’ goals, intentions, knowledge of the content, and knowledge of their students impact the way the task is set up in the lesson. Third, the classroom features and students’ dispositions impact the way the task is enacted.

Thus, the potential of a mathematical task for teaching can be seen as a combination of its inherent features and the teacher’s knowledge, beliefs, and goals, while its potential for individual students’ learning depends on students’ dispositions. Although many studies confirm this assumption of varying task potentials (Boston & Smith, 2011; Kullberg et al., 2014; Stein et al., 1996; Sullivan et al., 2009) and produce new theoretical reflections on the matter (e.g., a cyclical model of task design and implementation; Thanheiser, 2017), it remains clear that the task as represented in the material is the foundation for further task implementation (Boston & Wolf, 2006). Analyzing a task’s inherent potential is therefore of high interest for research related to mathematical tasks.

In terms of students’ mathematical learning gains, the focus is often on the tasks’ potential for cognitive activation. This construct has been conceptualized differently in previous studies, mostly either as a stand-alone dimension of the task or as the combination of separate different indicators. One of the first approaches to determine a task’s potential for cognitive activation in a holistic way was Bloom’s taxonomy of educational objectives (Bloom et al., 1956). While the original taxonomy was solely focused on cognitive activities, a revised version developed by Anderson and Krathwohl (2001) identified different knowledge facets (factual, conceptual, procedural, and metacognitive) as a second key dimension in addition to cognitive activities (remember, understand, apply, analyze, evaluate, and create). A similar, but mathematics-specific, approach to describing the cognitive demand of tasks was introduced by Stein et al. (1996). Their four levels of mathematical activities “range from memorization, to the use of procedures and algorithms [...], to the employment of complex thinking and reasoning strategies that would be typical of ‘doing mathematics’ (e.g., conjecturing, justifying, interpreting, etc.)” (p. 461). However, analyses, such as those in the context of the Instructional Quality Assessment (IQA), show that it is difficult for multiple raters to reliably assess tasks based on these four categories (Boston & Wolf, 2006). While this is a common challenge for high-inference ratings, it may also indicate that the cognitive demand of tasks is a complex, multi-dimensional construct, which is difficult to adequately capture using only one characteristic.

This holistic perspective on the potential for cognitive activation of tasks is contrasted in other studies with atomistic approaches in which different dimensions of cognitive activation are considered. In German-speaking countries, the Cognitive Activation in the Classroom Study (COACTIV) project has shaped discussions about the potential of mathematical tasks (Baumert et al., 2010; Neubrand et al., 2013). The classification scheme for task analysis (Jordan et al., 2006) originally included 18 categories and captured a wide variety of aspects. However, three dimensions—*type of mathematical task (purely technical, computational modelling, and conceptual modelling)*, *level of mathematical argumentation required*, and *translation processes within mathematics*—were deemed most suitable to describe tasks' potential for cognitive activation with regard to the quality of instruction (Baumert et al., 2010). However, to the best of our knowledge, there has been no in-depth analysis of the relations between the aforementioned dimensions and the way they determine a task's potential for cognitive activation.

Recent analyses in the context of the Global Teaching InSights Study (TALIS-Video, originally the Teaching and Learning International Survey Video Study) suggest that similar indicators for different aspects of the potential for cognitive activation can be integrated into a common construct, thus integrating the holistic and the atomistic approach. Herbert and Schweig (2021) established six categories for the analysis of classroom artifacts: *connecting mathematical representations*, *real-world contexts*, *asking for explanations*, *using multiple mathematical methods*, *encouraging self-evaluation*, and *linguistic complexity*. The results point out that the aforementioned variables represent a common construct. However, the study was limited to lessons on quadratic equations and functions.

Another approach to reach a comprehensive understanding of the task as a whole based on individual characteristics originates from the discussion about (cognitive) mathematical competencies. Tasks, as the main carriers of mathematical content and activities, need to be designed in a way that enables the development of various mathematical competencies, as this is one of the central goals of mathematics education (Niss & Højgaard, 2011). While there is no common definition of the term (*mathematical*) *competency* (Blömeke et al., 2015), competency frameworks have emerged as goals and guidelines for mathematics instruction (for an overview, see Pettersen & Nortvedt, 2018). One of the most renowned frameworks originated from the Competencies and the Learning of Mathematics Framework (KOM framework; Niss & Højgaard, 2011, 2019). In the adaptation for the Programme for International Student Assessment (PISA, Niss, 2015), Turner et al. (2015) describe how mathematical tasks can be characterized by the extent to which they foster each of the following six competencies: communication; devising strategies; mathematization; representation; using symbols, operations, and formal language; and reasoning and argument (for an overview, see Turner et al., 2023). It should be noted that the eight competencies in the KOM framework were explicitly designed to be *generic*, that is, “independent of specific mathematical subject matter as well as of specific educational levels” (Niss & Højgaard, 2019, p. 10), as well as “distinct, but not disjoint” (ibid., p. 19). Empirical analyses using PISA items confirm this second assumption for the six competencies included in Turner et al. (2015), revealing moderate positive correlations between mathematical competencies and showing that the levels of different competencies required to solve an item can be used to predict the difficulty of an empirical item (Pettersen & Braeken, 2019; Turner et al., 2013).

As outlined in this section, the potential for cognitive activation of tasks has so far been considered and assessed from different perspectives. A holistic approach classifies tasks either as challenging (cognitively activating) or as less challenging, without taking

a detailed look at individual characteristics. An atomistic approach often does not sufficiently clarify the extent to which different facets are related and how this relation shapes the tasks' potential for cognitive activation. Thus, the aim of the present work is to contribute to bridging the gap between the described holistic and atomistic approaches. To this end, different types of tasks are identified empirically based on the extent to which they foster different mathematical competencies (*sensu* Niss, 2015). These resulting types of tasks are then examined with regard to the levels of cognitive mathematical activities, departing from Anderson and Krathwohl (2001), as well as Stein et al. (1996), thus combining different approaches to determine the tasks' potential for cognitive activation. This exploratory approach hereby offers an opportunity to empirically validate and illustrate different ways in which tasks can elicit the various cognitive mathematical activities.

In this way, the present study connects to prior work on the potential of tasks by understanding tasks as an integral part of teaching. As stimuli for mathematical activities, tasks provide opportunities for students to demonstrate and develop general mathematical competencies; thus, the analysis of task features is seen as an important component for understanding teaching and learning processes in the classroom and their potential for fostering cognitive activation. For further frameworks that focus on task design as a process, incorporate the perspectives of other groups (e.g., students), or establish detailed theories for individual mathematical topics, we refer in particular to work from the International Commission on Mathematical Instruction group on task design (ICMI-22; for an overview, see Kieran et al., 2015).

### 3 Conceptual framework and research questions

The present study is situated in the Teacher Education and Development Study (Learning to Teach Mathematics, TEDS-M) research program within the TEDS-Validate study, which focuses on the relations between mathematics teachers' professional competencies and their students' learning gains (Kaiser et al., 2017). The initial research interest of TEDS-M was the analysis and comparison of teachers' mathematical (pedagogical) content knowledge (MCK/MPCK; Shulman, 1986). However, German follow-up studies have expanded the research on the underlying effect chain between teachers' competence and students' achievement to include teachers' situated skills (Blömeke et al., 2015) and the quality of instruction (Jentsch et al., 2021b; Schlesinger et al., 2018) as mediating variables (for an analysis of the full effect chain, see Blömeke et al., 2022). To validate the lesson observations and investigate instructional quality in more detail, the potential for cognitive activation of the learning tasks (PCAT) used during instruction was included in the model as part of instructional quality. Tasks, as learning opportunities, have a high PCAT if they lead to a deeper understanding of mathematical content and interconnections of concepts (Kunter & Voss, 2013). To this end, tasks should build on students' ways of thinking and prior knowledge, and they should support metacognition and higher-level thinking (Praetorius et al., 2018).

In accordance with prevalent works in the didactical discourse (e.g., Bromme et al., 1990; Doyle, 1983; Neubrand, 2002; Stein et al., 1996), a *task* in the context of this study is seen as a *prompt asking the students to formulate a product* as the result of dealing with a *specified mathematical situation*. Thus, if the mathematical situation changes significantly or the students are required to formulate sufficiently different products, two (or more) different (sub-)tasks are considered in the analysis. For instance, task A (Table 1) is seen as a

**Table 1** Examples for the organization of tasks in units of analysis

Task A (1 unit of analysis)	Task B (12 sub-tasks/units of analysis)			
Calculate mentally if possible	Calculate mentally if possible			
$3 \cdot (-5) = -15$	a)	b)	c)	d)
$2 \cdot (-5) = -10$	$(-8) \cdot 3$	$17 \cdot (-3)$	$9 \cdot (-41)$	$4 \cdot (-65)$
$1 \cdot (-5) = -5$	$(-4) \cdot (-3)$	$(-8) \cdot (-24)$	$(-4) \cdot 63$	$(-8) \cdot 125$
$0 \cdot (-5) = 0$	$9 \cdot (-14)$	$(-72) \cdot 8$	$(-5) \cdot (-36)$	$4 \cdot (-250)$
$(-1) \cdot (-5) = \dots$				
$(-2) \cdot (-5) = \dots$				
$(\dots) \cdot (-5) = \dots$				

coherent entity because, in contrast to the twelve isolated sub-tasks of the second example from the same lesson, the focus is on recognizing the general pattern. In addition to the task wording, the solution paths available to the students are understood to be inextricably linked to the tasks and are thus part of the unit of analysis (Doyle, 1983).

The underlying assumption of this analysis is that high and low levels of PCAT can be attained and different levels of mathematical competencies can be promoted regardless of students' age group and the area of mathematical content addressed by the task. Departing from prior work on the analysis of mathematical tasks, particularly the COACTIV study (Jordan et al., 2006; Neubrand et al., 2013) and work connected to the Trends in International Mathematics and Science Study (TIMSS, originally the Third International Mathematics and Science Study; Neubrand, 2002), the following task characteristics are assumed to impact PCAT:

- Content-related characteristics (such as the curricular level or the interconnectedness of different content areas);
- The levels of different mathematical competencies (mathematical modelling, problem solving, reasoning and argumentation, the use of representations, the use of symbols and operations, and mathematical communication) required to solve the task (following Turner et al., 2015);
- The cognitive complexity of the task (departing from Anderson & Krathwohl, 2001);
- The linguistic complexity of the task formulation (using indicators for complexity that are specific to the German language); and
- Overarching task characteristics, such as closeness to reality and openness.

Following Niss and Højgaard (2019) in their notion of *distinct, but not disjoint* competencies, the focus is on the characteristic features of the respective competency, that is, on the “well-defined identity which singles it out from the other competencies” (p. 19). Therefore, for mathematical modelling, the active transitions between mathematics and the real world are examined, while devising and applying strategies for the solution of mathematical problems is seen as a key component of problem solving. For the reasoning and argumentation competency, both the comprehension and the production of (chains of) arguments are seen as relevant. The isolated handling of symbolic and verbal representations is already emphasized with regard to the use of symbols and operations, respectively, the mathematical communication competency. Thus, in the use of mathematical representations, especially the iconic and enactive level as well as the deliberate selection of mathematical representations and the transitions between

different forms of representation are considered. Following the work of Turner et al. (2015) in the context of the PISA tasks, both the mathematical thinking competency and the mathematical aids and tools competency were not included in our task analysis.

Previous analyses suggest that different mathematical competencies are not always addressed to the same degree (Neubrand et al., 2013; Pettersen & Braeken, 2019; Turner et al., 2015). We therefore assume that different types of tasks require the aforementioned mathematical competencies to varying degrees, and thus, cognitive activation can be promoted in different ways. In a first step, we aim to investigate the relations between these different task characteristics and identify patterns in the interplay of the mathematical competencies within the tasks sampled from German classrooms. Our first research question (RQ) is as follows:

(RQ1) What are the different types of tasks that can be identified within the sample of instructional tasks from various lessons based on their potential to foster different mathematical competencies?

Following the previous approaches to task analysis outlined in Section 2, PCAT cannot be fully understood by focusing only on mathematical competencies. The revised taxonomy developed by Anderson and Krathwohl (2001), for instance, emphasizes the need to consider the knowledge required to solve a task in addition to (mathematical) activities. Therefore, in order to gain a more detailed insight into the potential for cognitive activation of the different task types (RQ1), the correlation with the type of knowledge and cognitive mathematical activity required to solve the tasks is analyzed.

Within this frame, the second research question has been formulated as follows:

(RQ2) To what extent do the resulting types of tasks (RQ1) differ with regard to the type of knowledge and cognitive mathematical activity required to solve the task?

## 4 Methodology

### 4.1 Study design and sample

The data stems from lower secondary mathematics classrooms in three different German federal states (Hesse, Saxonia, and Thuringia).<sup>2</sup> A total of 38 teachers volunteered to participate in all parts of the TEDS-Validate study, including tests of teachers' knowledge and situation-specific skills related to noticing, a questionnaire on their beliefs concerning mathematics education, and two lesson observations (Kaiser et al., 2017). For the latter, two trained raters assessed different aspects of instructional quality in vivo (Jentsch et al., 2021a; Schlesinger et al., 2018). For 31 of the aforementioned participating teachers, all tasks set throughout the course of the lesson were sampled by writing down oral assignments (9.8%) and writings on the blackboard (12.2%), as well as gathering pictures of the textbooks (32.7%) and worksheets (45.2%). Both the grade level (5th to 10th grades) and the mathematical content of lessons varied widely between the teachers as they decided on their own, in whichever classes observations were possible. The study included 60 lesson

<sup>2</sup> Due to the federalism of education in Germany, all federal states have their own educational systems and state curricula.

observations,<sup>3</sup> mostly with a duration of 90 min each. The number of tasks per teacher ranged from 15 to 197, resulting in an overall sample of 2490 tasks. To ensure high levels of rater agreement and accuracy in the assessment of task potential, four pre-service teachers drafted one or more possible solution paths for each of the tasks and marked the most likely path based on careful analysis of students' expected prior knowledge and competencies based on the state curricula for each grade level. These solution paths were then checked and, if necessary, revised by the first two authors.

## 4.2 Task classification in TEDS-Validate

Regarding the scope of task analysis, Resnick (1975) distinguishes between rational and empirical task analyses. This distinction aligns with the different phases of task implementation proposed by Stein et al. (1996). Rational task analysis is based on the wording of the task itself as well as possible successful approaches to solving the task, thus focusing on its inherent potential. Empirical task analysis, in contrast, includes actual solution paths based on transcripts of students' notes or self-reports. This type of analysis is better suited to assess the realized potential of the task. Since the focus of this paper is tasks and their PCAT, a rational task analysis was conducted.

As outlined in the previous sections, a comprehensive analysis of the tasks' potential requires consideration of various task characteristics. In the context of TEDS-Validate, a classification scheme for rational task analysis that includes different dimensions was developed based on prior work, mainly the COACTIV study (Jordan et al., 2006) and the PISA framework (Turner et al., 2015). The classification scheme includes high-inference rating scales to examine tasks with regard to surface features (e.g., the mathematical content area(s) and target grade level), underlying mathematical concepts and ideas, mathematical competencies, and cognitive and linguistic complexity. While surface features do not necessarily allow for consistent scaling, all other characteristics were assessed using a 4-point ordinal scale (0–3). Following the idea of *generic* competencies—in the sense that high and low levels of competency can be attained regardless of the grade level and the mathematical content considered—the analyses were conducted with the expected knowledge levels and competencies of the students in mind. In the electronic supplementary material, the scale used for the modelling category is presented as an example. Similarly, the ratings for the other competencies are based on the characterization of complexity, abstractness, and independent mathematical thinking in terms of their respective *well-defined identity* following Niss and Højgaard (2019; see Section 3).

The analyses presented in this paper are based on the following dimensions:

- The potential of each task with regard to modelling, problem solving, reasoning and argumentation, use of representations, symbols and operations, and communication, each assessed on an ordinal scale from 0 to 3 (no potential to high potential);
- The knowledge facet predominantly required for the solution of the task (factual, procedural, conceptual, or metacognitive knowledge), assessed as a nominal variable;
- Four levels of cognitive mathematical activities (remember/reproduce—understand/apply—analyze/evaluate—create).

<sup>3</sup> For one of the teachers, only one lesson could be observed due to illness. In another case, not all tasks used in the lesson could be sampled and thus the lesson had to be excluded.



Pre-service mathematics teachers were trained as raters until satisfactory rater agreement was reached. Twenty percent of the sample ( $n=2490$ ) was double coded. The two raters then discussed any disagreements and adjusted the final coding of each item if necessary. Overall agreement for the different categories was satisfactory to very good ( $0.662 \leq \kappa \leq 0.974$ ).

As shown in Table 2, the majority of tasks offer little potential for the use of the mathematical competencies. Very few tasks require a high level of one or more of the six competencies; a need for reasoning and argumentation is apparent in less than 5% of the tasks. These results are consistent with previous studies on mathematical tasks in German-speaking countries, having shown an alarmingly low potential for the promotion of these competencies and an exclusive focus on calculations and the technical elements of mathematics (Brunner et al., 2019; Driike-Noe, 2014; Neubrand, 2002; Neubrand et al., 2013). Despite significant reform movements concerning German mathematics teaching, the quality of mathematical tasks has apparently not been significantly improved. In order to verify that the operationalizations of the competencies reflect sufficiently diverse aspects of the tasks' potential, Kendall's rank correlation coefficient suitable for ordinal variables was considered. The lack of moderate or high correlations between the different mathematical competencies (see Table 3) indicates that they can indeed be clearly distinguished.

**Table 2** Levels of different competencies required by the tasks

Domain	Code 0	Code 1	Code 2	Code 3
MOD	2239 (89.9%)	190 (7.6%)	60 (2.4%)	1 (0%)
PROB	1957 (78.6%)	433 (17.4%)	88 (3.5%)	11 (0.4%)
R&A	2403 (96.5%)	69 (2.8%)	14 (0.6%)	4 (0.2%)
REP	1396 (56.1%)	765 (30.7%)	296 (11.9%)	33 (1.3%)
S&O	864 (34.7%)	1280 (51.4%)	304 (12.2%)	42 (1.7%)
COM	1355 (54.4%)	791 (31.8%)	277 (11.1%)	67 (2.7%)

*MOD*, mathematical modelling; *PROB*, problem solving; *R&A*, reasoning and argumentation; *REP*, use of representations; *S&O*, symbols and operations; *COM*, communication

**Table 3** Correlations between different mathematical competencies

Domain	MOD	PROB	R&A	REP	S&O	COM
MOD	1	.043*	-.012	.149***	-.078***	.294***
PROB		1	.144***	.077***	.101***	.254***
R&A			1	-.035	-.031	.215***
REP				1	-.288***	.226***
S&O					1	-.299***
COM						1

*MOD*, mathematical modelling; *PROB*, problem solving; *R&A*, reasoning and argumentation; *REP*, use of representations; *S&O*, symbols and operations; *COM*, communication. The table shows values for Kendall's tau-b, with  $0.1 < \tau < 0.3$  considered to indicate a weak effect,  $0.3 < \tau < 0.5$  considered to indicate a medium effect, and  $0.5 < \tau$  considered to indicate a strong effect (Cohen, 1988). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

### 4.3 Data analysis

Due to the categorical nature of the independent variables, a latent class analysis (LCA) was carried out using the software Mplus 7.4 (Muthén & Muthén, 1998-2015). The LCA identified types of mathematical tasks that require a similar set of competencies. The assumption underlying LCA is the existence of a latent categorical variable (the *classes*) that explains patterns observed in the data. Each of the objects in the analysis has a certain probability of belonging to each of the different classes (Vermunt & Magidson, 2002; Weller et al., 2020). This method was chosen over deterministic approaches to clustering because the number of classes in this exploratory analysis was a priori unclear<sup>4</sup> and the description of task prototypes was of higher interest than accurate allocation of individual tasks to a class.

Due to the small number of tasks at the upper end of the scale, codes 2 and 3 were aggregated following the principle of parsimony (Epstein, 1984). This resulted in 3-point ordinal scales for all six competencies. *Reasoning and argumentation* was excluded as an indicator variable because it occurred too seldom in the data set and thus increased the model's complexity without contributing sufficient information. As the correlations between the remaining five indicator variables in the overall data set were low, it was presumed that the local independence assumption, a prerequisite for LCA, could be upheld.

The number of latent classes was determined based on different information criteria: Akaike's information criterion (AIC; Akaike, 1987), the Bayesian information criterion (BIC), and its sample-size-adjusted version (ABIC; Schwartz, 1978). However, these different criteria do not always point to the same model and can tend to overfit or underfit the true model (Dziak et al., 2020; Nylund et al., 2007). Thus, the AIC, BIC, and ABIC were only used to narrow down the eligible models. Then, the remaining models were evaluated based on their entropy, which served as a measure for uncertainty (Ramaswamy et al., 1993), the size of the smallest class, and, most importantly, the theoretical interpretability. This process led to the selection of a final model. A priori, a model of ten or more classes (double the number of indicator variables) was considered overly complex. Thus, the number of classes was limited to nine or less.

For further post hoc analyses, all tasks were assigned to the class to which they were most likely to belong based on the estimated probabilities. To investigate the distribution of the knowledge facets and the cognitive mathematical activities among the different classes, a chi-square test, a Kruskal–Wallis test (Kruskal & Wallis, 1952), and a pairwise group comparison performed with a Mann–Whitney *U* test (Mann & Whitney, 1947), were carried out respectively using SPSS 29. As one of the mathematical competencies, *reasoning and argumentation*, could not be included as an indicator variable in the LCA, it was included in the post hoc analyses.

## 5 Results

### 5.1 Number of latent classes

The analysis with regard to the first research question displayed in Table 4 points to a six-class model.

<sup>4</sup> This also holds true for forms of hierarchical cluster analysis, but these are less suitable for bigger sample sizes, such as the one in the present study ( $n = 2490$ ).

**Table 4** Information criteria for the selection of a model in LCA

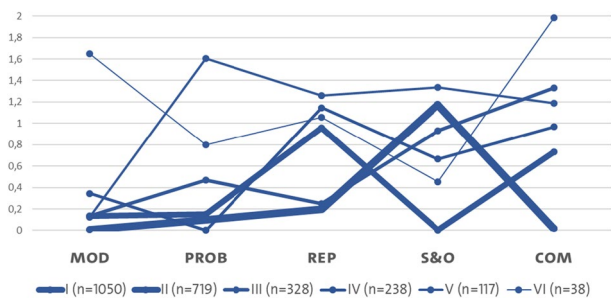
#Classes	AIC	BIC	ABIC	Entropy	$p_{\min}$	$n_{\min}$
2	18,453.451	18,575.671	18,508.949	0.755	0.892	1005
3	18,235.390	18,421.632	18,319.960	0.768	0.773	266
4	18,096.266	18,346.528	18,319.960	0.823	0.861	124
5	18,013.174	18,327.456	18,155.885	0.781	0.748	82
<b>6</b>	<b>17,940.382</b>	<b>18,318.685</b>	<b>18,112.164</b>	<b>0.841</b>	<b>0.772</b>	<b>38</b>
7	17,901.467	18,343.790	18,102.319	0.751	0.733	33
8	17,875.568	18,381.911	18,105.490	0.813	0.729	28
9	17,865.669	18,436.033	18,124.662	0.817	0.701	33

*AIC*, Akaike's information criterion; *BIC*, Bayesian information criterion; *ABIC*, sample-size adjusted Bayesian information criterion. For all three, lower values indicate better model fit. Bold values show the values for the six-class model, which was ultimately chosen.  $p_{\min}$  = minimum average class probability,  $n_{\min}$  = number of tasks in the smallest resulting class

The BIC and ABIC indicate the best model fit for six and seven classes, respectively. The AIC did not assume a minimal value for up to nine classes and was excluded from consideration due to its tendency to overfit (Dziak et al., 2020). Out of the six-, seven-, and eight-class solutions suggested by the remaining information criteria, the six-class solution showed slightly better values for entropy, class probabilities, and minimum class size. Furthermore, the six classes revealed interesting qualitative differences, while the addition of more classes—which, in some instances, only differed quantitatively—hindered the interpretability of the overall model. Thus, the six-class solution was selected for this study.

## 5.2 Types of mathematical tasks in the sample

The resulting classes (RQ1) show varying levels of potential for fostering the five mathematical competencies that served as indicator variables (see Fig. 1). In order to illustrate the differences between the classes, sample tasks from the classroom observations, which are typical for each class in relation to the distribution of competencies, are shown in Table 5.



**Fig. 1** Distribution of mathematical competencies across different classes.

*Note.* *MOD*, mathematical modelling; *PROB*, problem solving; *REP*, use of representations; *S&O*, symbols and operations; *COM*, communication. Left axis indicates the mean values (ordinal coding, levels 0–2) for each class

**Table 5** Sample tasks for each of the six classes

<b>Class I (n = 1050)</b>			<p>(1) <math>2.5m^2 = \underline{\hspace{2cm}} cm^2</math></p> <p>(2) <math>x^2 + 20x + 100 = (\dots)^2</math></p>						
Coding for grade 7 (1) and grade 8 (2)	MOD	0							
	PROB	0							
	REP	0							
	S&O	1							
	COM	0							
<b>Class II (n = 719)</b>			<p>Which of the labels of the triangles does not correspond to the given equation?</p> <p>a) <math>\sin \alpha = \frac{a}{b}</math>      b) <math>\frac{u}{\sin \alpha} = \frac{v}{\sin \beta}</math></p> <p>c) <math>r^2 = t^2 - s^2</math>    d) <math>x^2 = y^2 + z^2 - 2yz \cdot \cos \alpha</math></p>						
Coding for grade 10	MOD	0							
	PROB	0							
	REP	1							
	S&O	0							
	COM	1							
<b>Class III (n = 328)</b>			<p>For which value of x is the term <math>\frac{4x-3}{6x+6}</math> undefined?</p>						
Coding for grade 10	MOD	0							
	PROB	1							
	REP	0							
	S&O	1							
	COM	1							
<b>Class IV (n = 238)</b>			<p>Read the temperature change on the thermometer and give the new temperature.</p> <table border="1" style="display: inline-table; margin-right: 20px;"> <tr> <td>Old temperature</td> <td>+3°C</td> </tr> <tr> <td>Temperature change</td> <td></td> </tr> <tr> <td>New temperature</td> <td></td> </tr> </table>	Old temperature	+3°C	Temperature change		New temperature	
Old temperature	+3°C								
Temperature change									
New temperature									
Coding for grade 7	MOD	1							
	PROB	0							
	REP	1							
	S&O	1							
	COM	1							
<b>Class V (n = 117)</b>			<p>What is the equation of the straight line to which the function given by <math>y = x^2 - 8x + 3</math> is axisymmetric?</p>						
Coding for grade 9	MOD	0							
	PROB	2							
	REP	2							
	S&O	1							
	COM	2							
<b>Class VI (n = 38)</b>			<p>In pairs, invent a “bathtub story.” Draw a diagram indicating the filling level of a bathtub (like the one below) and write down the corresponding story of a person taking a bath.</p>						
Coding for grade 8	MOD	2							
	PROB	1							
	REP	2							
	S&O	0							
	COM	2							

*MOD*, mathematical modelling; *PROB*, problem solving; *REP*, use of representations; *S&O*, symbols and operations; *COM*, communication. The ordinal ratings for each task can be seen as typical for tasks from the respective class

The largest class, which included nearly half of the tasks, can be characterized as calculation oriented due to the exclusive focus on the use of symbols and operations. None of the tasks from the second-largest class shows a potential for this competency. Instead, they tend to require the skilled use of representations and/or the reception and production of mathematical text. Active and passive engagements with mathematical texts also play a major role in solving tasks from the third class ( $n=328$ ); however, the tasks also require the use of symbols and formalism, and low-level problem solving skills. While the fourth class shows moderate potential for several of the analyzed competencies, interestingly, none of the tasks require any problem solving. This indicates that these tasks are designed to practice and repeat well-known mathematical procedures. Tasks from the fifth class ( $n=117$ ) require moderate to high levels of all competencies except mathematical modelling. The last and smallest class ( $n=38$ ) is the only one to show a high potential for mathematical modelling while also incorporating mathematical communication on a higher level than all other classes.

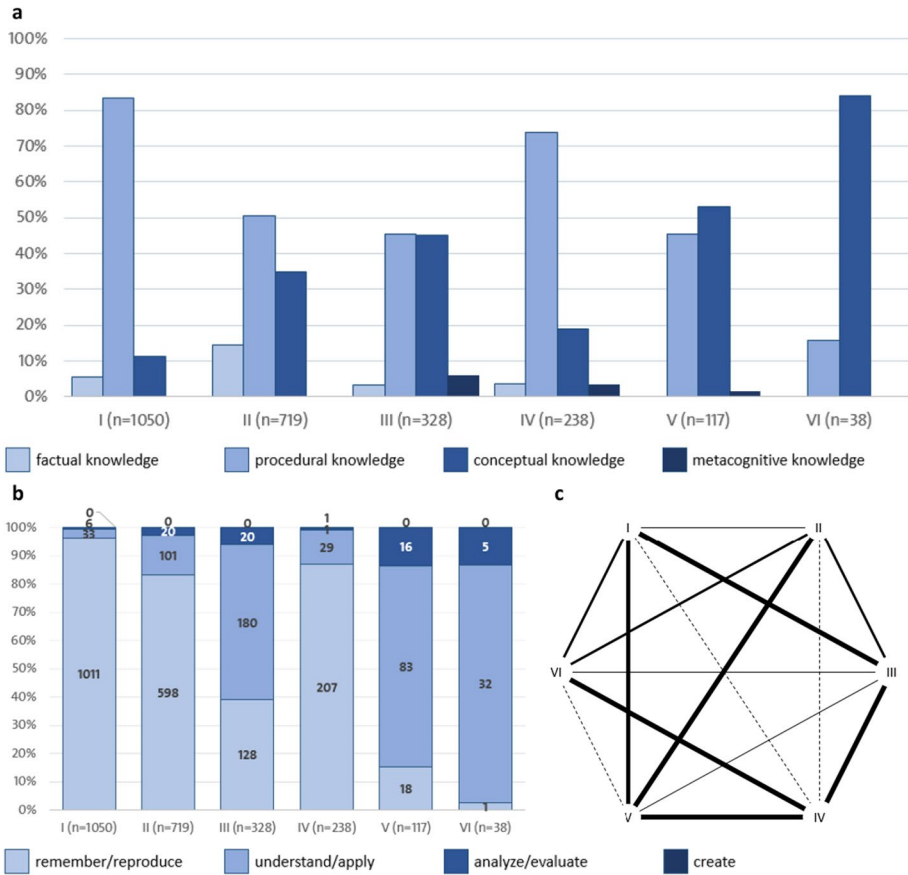
### 5.2.1 Knowledge facets and cognitive mathematical activities across the different classes (RQ2)

To gain a deeper understanding of the task classes presented above, the following section highlights the extent to which they differ with regard to the knowledge facet (factual, procedural, conceptual, and metacognitive) as well as the cognitive mathematical activity (remember/reproduce—understand/apply—analyze/evaluate—create) predominantly required in the solution of the tasks. Analyses were performed post hoc on the task classes described in Section 5.2. A chi-square test shows that the six classes of tasks differ significantly with respect to the dominant knowledge facet,  $\chi^2(15, 2490)=527.075$ .<sup>5</sup> Estimates of the Kruskal–Wallis test reveal statistically significant differences for the cognitive mathematical activities ( $H=927.392$ ,  $p<0.001$ ,  $df=5$ <sup>6</sup>). Figure 2 shows the descriptive results across the six classes and the results of pairwise comparison of the classes, which was performed with the Mann–Whitney  $U$  test. The average ranks for cognitive mathematical activities resulting from the Kruskal–Wallis test are illustrated in Table 6.

It first becomes apparent that the tasks from the sample require predominantly procedural and conceptual knowledge on the one hand and the activities of reproducing and applying knowledge on the other. Tasks with a focus on the use of metacognitive knowledge and more complex cognitive activities like evaluating or creating are rare to non-existent. The results also reveal that differences in the types of knowledge primarily addressed tend to be gradual across classes, with classes I and IV, and class VI at both ends of the spectrum. However, two subgroups of classes emerge when considering the

<sup>5</sup> It should be noted that, due in particular to the rare occurrence of metacognitive knowledge, one-fifth of the cells had an expected frequency of less than 5. Due to the clear result and the large effect size ( $d=1.04$ ), however, it seemed justifiable to forego Fisher's exact test, which would require a disproportionately large amount of computational effort, given the size of the sample.

<sup>6</sup> For the *reasoning and argumentation* competency, which had to be excluded as an indicator variable for the LCA due to the lack of variance, the Kruskal–Wallis test reveals statistically significant differences ( $H=221.393$ ,  $p<.001$ ,  $df=5$ ). However, an in-depth analysis of the distribution across different classes confirms low variance with regard to the necessity for *reasoning and argumentation*. With the exception of a few tasks in the other classes, only class III contains a notable number of tasks that encourage students to justify their approaches and results or develop more advanced mathematical reasoning. Mann–Whitney  $U$  tests show that the only significant (but small) pairwise differences with regard to *reasoning and argumentation* can be observed between class III and tasks from other classes.



**Fig. 2** Distribution across and differences between different classes for knowledge facets (a) and cognitive mathematical activities (b). *Note.* a and b The horizontal axis indicates the different classes. The vertical axis indicates the percentages of each knowledge facet and cognitive mathematical activity. (c) --- No significant ( $p \geq 0.05$ ) or no effect ( $d < 0.2$ ); — small effect ( $0.2 < d < 0.5$ ); — intermediate effect ( $0.5 < d < 0.8$ ); — large effect ( $d \geq 0.8$ ) (Cohen, 1988)

**Table 6** Distribution of cognitive mathematical activities across classes

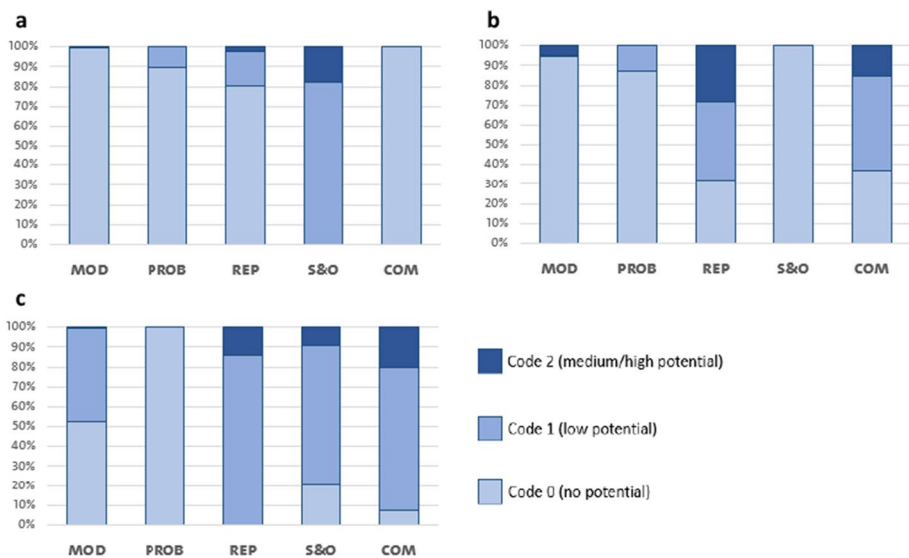
	<i>n</i>	Remember/reproduce	Understand/apply	Analyze/evaluate	Create	Average rank
Class I	1050	1011 (96.3%)	33 (3.1%)	6 (0.6%)	0 (0%)	1461.54
Class II	719	598 (83.2%)	101 (14%)	20 (2.8%)	0 (0%)	1296.97
Class III	328	128 (39%)	180 (54.9%)	20 (6.1%)	0 (0%)	753.85
Class IV	238	207 (87%)	29 (12.2%)	1 (0.4%)	1 (0.4%)	1347.97
Class V	117	18 (15.4%)	83 (71%)	16 (13.7%)	0 (0%)	447.76
Class VI	38	1 (2.6%)	32 (84.2%)	5 (13.2%)	0 (0%)	294.75

The average ranks for tasks from each class are calculated compared to the entire sample ( $n = 2490$ )

predominant cognitive mathematical activities—classes I, II, and IV, and classes III, V, and VI (see Fig. 2c and Table 6). While tasks from the first three classes focus mainly on remembering and reproducing, thus cognitive mathematical activities of a lower cognitive complexity, the majority of tasks from the other classes require understanding and application and—in some cases—leave room for higher-order thinking. The two subgroups shall be described in more detail in the following sections.

### 5.2.2 Types of routine tasks in the sample

These first three classes contain with just over 2000 tasks about 80% of the entire data set. They are summarized in the following under *routine tasks*, as they generally focus on reproduction rather than on more complex cognitive activities (see Section 5.2.1) and require almost no problem solving comprising more than one-step solutions. However, the individual classes differ in terms of the mathematical competencies necessary to solve the respective tasks (Fig. 3). The largest class focuses exclusively on the use of symbols and operations while other competencies are only found in exceptional cases and at a low level. In contrast, none of the tasks from the second class requires (proficient) usage of symbols and operations. Instead, other forms of representation, i.e., graphic and/or verbal, seem to be of high importance in this class. Hence, these two types of routine tasks can be characterized as *calculation-oriented* and *representation-oriented routine tasks*. In contrast to the other two classes, all tasks from class IV require the use of multiple competencies and, possibly, their integration in the solution process. Interestingly, among these *composite routine tasks* (class IV), as many entail purely intra-mathematical engagement as they do active engagement with an extra-mathematical context.

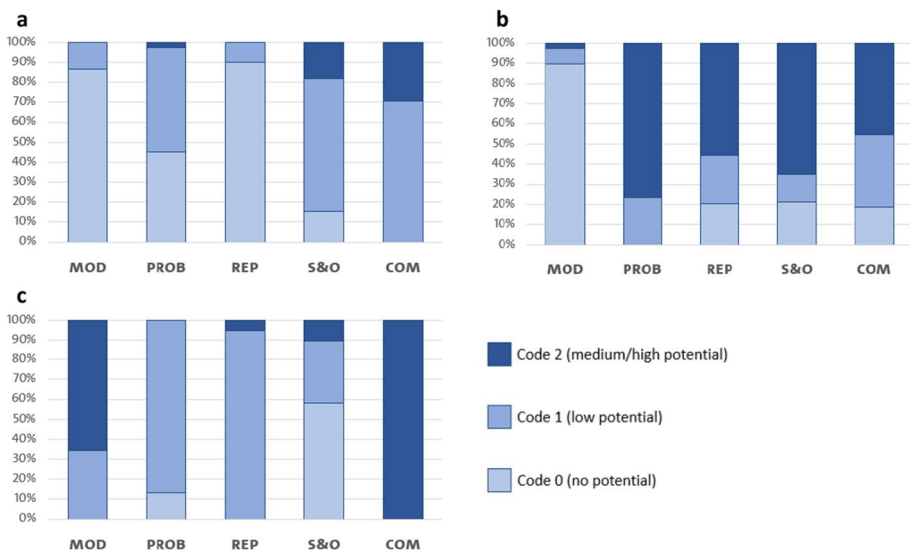


**Fig. 3** Distribution of competencies among *routine tasks*. **a** Class I ( $n=1050$ ), **b** class II ( $n=719$ ), **c** class IV ( $n=238$ ).

*Note.* MOD, mathematical modelling; PROB, problem solving; REP, use of representations; S&O, symbols and operations; COM, communication. Vertical axis indicates the percentage of tasks at each level for the different competencies

### 5.2.3 Application tasks

The majority of tasks from the classes presented in this section show a focus on the application of mathematical concepts and methods in new (mostly mathematical) contexts rather than the mere reproduction of knowledge (see Section 5.2.1). This tendency is also reflected in the need for varying levels of problem-solving skills as well as—in the case of classes V and VI—an overall higher potential for fostering several of the competencies considered in this analysis. The nature and complexity of these more or less unfamiliar contexts, however, seem to vary between the three types of *application tasks* (Fig. 4). The challenge in solving tasks from class III appears to lie mainly in understanding and producing mathematical texts and choosing an appropriate solution method based on familiar symbols and operations, both on a lower level of complexity. Thus, these tasks can be characterized as *simple word problems*. In contrast, tasks from class V typically foster higher levels of different competencies, with a specific focus on problem solving. Most of the tasks do not require consideration of a real-world context; rather, they focus on inner-mathematical contexts. Therefore, the tasks in this class, which account for approximately 5% of the sample, can be described as (simple) *inner-mathematical problems*. The smallest class of tasks ( $n=38$ ) can be characterized by a focus on mathematization and interpretation activities, while also requiring many other competencies—albeit to a lower extent. As can be expected for tasks with rich real-world contexts, the challenge of understanding potentially difficult contextual descriptions and recognizing relations within and beyond the task formulation is reflected in the high ratings for the mathematical communication competency. However, these *real-world problems* represent only 1.5% of the total sample.



**Fig. 4** Distribution of competencies among *application tasks*. **a** Class III ( $n=328$ ), **b** class V ( $n=117$ ), **c** class VI ( $n=38$ ).

*Note.* MOD, mathematical modelling; PROB, problem solving; REP, use of representations; S&O, symbols and operations; COM, communication. Vertical axis indicates the percentage of tasks at each level for the different competencies



## 6 Summary and discussion

In the context of this study, tasks are seen as stimuli for mathematical (learning) activities with varying potential for cognitive activation. Following prior work in the field, aspects such as the interplay of knowledge facets and cognitive (mathematical) activities (Anderson & Krathwohl, 2001; Stein et al., 1996) as well as the need for mathematical competencies (Niss & Højgaard, 2019) are assumed to influence and indicate different levels of PCAT. The purpose of the present study was to identify different types of tasks based on the mathematical competencies required for their solution. To this end, a rational task analysis was performed on 2490 mathematical tasks taken from 60 lesson observations in Germany. LCA was carried out to identify six distinct classes of tasks within the sample, which show different levels of potential for developing the mathematical competencies (modelling, problem solving, use of representations, symbols and operations, and communication—reasoning and argumentation was excluded as an indicator due to its low empirical occurrence). Further analyses revealed that the six classes vary with regard to the knowledge facets and cognitive mathematical activities required.

The overall PCAT for fostering competencies is very low, with only few tasks facilitating mathematization (10%) or reasoning activities (4%). Instead, most tasks have a technical focus on the execution of calculations and mathematical procedures. The results of the study are in line with previous studies on mathematics education in Germany, which reported similar findings (Herbert & Schweig, 2021; Neubrand, 2002; Neubrand et al., 2013), suggesting that these observations seem to be stable concerning the quality of tasks in mathematics education in the last decades. Thus, although a clear focus on developing different mathematical competencies was established in German national curricula (Prenzel et al., 2015), this ideal has not significantly changed the quality of tasks in German classrooms.

An in-depth analysis of the resulting task classes in relation to the predominant cognitive mathematical activities revealed three types of *routine tasks* requiring mainly the reproduction of knowledge, and three types of tasks with a focus on different kinds of *application of mathematical knowledge*. A common characteristic within the routine tasks, which account for about 80% of the total sample, was the lack of potential for problem solving and devising (own) strategies. While the two largest classes showed an isolated focus on either symbolic ( $n=1050$ ) or graphical and/or verbal representations ( $n=719$ ), *composite routine tasks* ( $n=238$ ) combine these different competencies on a low level. The three classes focusing mainly on application and—in some cases—higher-order cognitive activities all require problem solving skills to some extent in combination with other mathematical competencies. The resulting classes of tasks can be characterized as *simple word problems* ( $n=328$ ), *inner-mathematical problems* ( $n=117$ ), and *real-world problems* ( $n=38$ ). Within the given sample, no types of tasks with a focus on more complex cognitive activities such as analysis and creation could be identified. With regard to the knowledge facets, a similar trend is visible—albeit with more gradual differences between the identified types of tasks—with the vast majority of tasks requiring predominantly procedural or conceptual as opposed to purely factual or metacognitive knowledge.

The possibilities that arise from a more in-depth perspective on tasks for research and practice are manifold. Considering different types of tasks based on the levels of mathematical competencies they require can lead to a deeper understanding of item demand in the context of student assessments. In the case of large-scale assessments, joint consideration of task features may lead to better prediction of item difficulty than individual characteristics. Analysis of the ratio of different task types can also be used to examine the comparability of final examinations in different districts, states, or countries. Additionally, analysis of the distribution of task types

as outlined in this paper can help paint a clearer picture of mathematics instruction throughout the course of a lesson than could analysis of isolated task features. This opportunity is particularly important when no further data on the quality of teaching can be collected. To fully realize these opportunities in research, further work is needed to shed light on the relationships between the tasks used and the quality of instruction.

Finally, the classification of tasks presented in this paper can be used for teachers' professional development as a means of bringing curricular reforms into classroom practice, since it can help to illustrate the more abstract concept of cognitive (mathematical) activities by means of the well-known mathematical competencies. In addition, the process of classification enables reflection on the interaction of individual competency demands required to solve tasks, thus developing teachers' mathematical task knowledge for teaching (for a first step, see Ross & Adleff, 2022).

Especially with regard to the opportunities highlighted above, the limitations of the study need to be carefully considered. All teachers participating in the TEDS-Validate study volunteered to take part in the time-consuming data collection. This resulted in a convenience sample of teachers with above-average commitment. It is therefore likely that the sample has positive bias. This is why the observed low PCAT seems even more significant.

Since the teachers were neither given any specifications about the subject area to be covered and the tasks to be used nor about the focus of the lesson (e.g., introduction of new content or practice and consolidation), the sub-samples from the lessons differed greatly in some cases. While the analysis instrument was specifically designed to be applicable across subject areas and grade levels, the strict definition of individual tasks as units of analysis led to varying numbers of tasks per lesson and teacher. Students typically worked on either a few complex tasks or many smaller, less time-consuming tasks in the same amount of time. Thus, the latter type of task is overrepresented in the sample. Use of weights in the LCA, which take into account the number of tasks in every lesson segment, did not yield significantly different results. Thus, to ensure the simplicity of the model, no weighting was applied. However, it is possible that the imbalance in favor of smaller, less time-consuming tasks may have impacted the resulting classes. Furthermore, while the assumption of distinctive features for each of the mathematical competencies (Niss & Højgaard, 2019) is endorsed by the lack of high correlations in-between, an influence on the composition of the classes cannot be entirely dismissed.

When drawing further conclusions about the quality of instruction in lessons, one must also consider that rational task analysis as well as the sole focus on PCAT and the indicators chosen for this study provide a limited perspective on what is happening in the classroom. To fully comprehend the complex reality of mathematics teaching and learning, additional information about the use of the tasks by both the teacher and individual students needs to be collected and linked to inherent task features.

Finally, it is important to emphasize that the analyses conducted in this study are partly exploratory in nature such as the use of LCA. Other contexts may lead to different classes of tasks, especially as the developed and used frameworks and instruments for task analysis may not be applicable in the same way to different cultural or educational settings. At the same time, considering different theoretical and empirical perspectives on tasks in mathematics education as well as cultural influences makes it necessary to contextualize the findings from this study and provide guidance for their use in practice.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data Availability** The data are available from the authors upon request.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**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

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317–332. <https://doi.org/10.1007/bf02294359>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. Longman.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education*, 56(3), 205–213. <https://doi.org/10.1177/0022487105275904>
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift Für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, 79, 101600. <https://doi.org/10.1016/j.learninstruc.2022.101600>
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. David McKay.
- Boston, M. (2012). Assessing instructional quality in mathematics. *The Elementary School Journal*, 113(1), 76–104. <https://doi.org/10.1086/666387>
- Boston, M. D., & Smith, M. S. (2011). A 'task-centric approach' to professional development: enhancing and sustaining mathematics teachers' ability to implement cognitively challenging mathematical tasks. *ZDM – Mathematics Education*, 43(6–7), 965–977. <https://doi.org/10.1007/s11858-011-0353-2>
- Boston, M., & Wolf, M. K. (2006). *Assessing academic rigor in mathematics instruction: The development of the Instructional Quality Assessment Toolkit*. National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Report #672. <https://doi.org/10.1037/e644922011-001>
- Bromme, R., Seeger, F., & Steinbring H. (1990). *Aufgaben als Anforderungen an Lehrer und Schüler* [Tasks as demands on teachers and students]. Aulis.
- Brunner, E., Jullier, R., & Lampart, J. (2019). Aufgabenangebot zum mathematischen Begründen in je zwei aktuellen Mathematikbüchern [Set of tasks for mathematical reasoning in two recent mathematics textbooks]. *Schweizerische Zeitschrift für Bildungswissenschaften*, 41(3), 647–664. <https://doi.org/10.24452/sjer.41.3.6>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Ed.)*. Routledge.
- Doyle, W. (1983). Academic work. *Review of Educational Research*, 53(2), 159–199. <https://doi.org/10.3102/00346543053002159>
- Drüke-Noe, C. (2014). *Aufgabenkultur in Klassenarbeiten im Fach Mathematik. Empirische Untersuchungen in neunten und zehnten Klassen* [Task culture in exams in mathematics. Empirical studies in ninth and tenth grade]. Springer Spektrum. <https://doi.org/10.1007/978-3-658-05351-2>
- Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefing in Bioinformatics*, 21(2), 553–565. <https://doi.org/10.1093/bib/bbz016>
- English, L. D. (2011). Complex learning through cognitively demanding tasks. *The Mathematics Enthusiast*, 8(3), 483–506. <https://doi.org/10.54870/1551-3440.1227>
- Epstein, R. (1984). The principle of parsimony and some applications in psychology. *The Journal of Mind and Behavior*, 5(2), 119–130.

- Heinle, A., Schiepe-Tiska, A., Reinhold, F., Heine, J. H., & Holzberger, D. (2022). Supporting student motivation in class: The motivational potential of tasks. *Zeitschrift Für Erziehungswissenschaft*, 25(2), 453–470. <https://doi.org/10.1007/s11618-022-01090-3>
- Herbert, B., & Schweig, J. (2021). Erfassung des Potenzials zur kognitiven Aktivierung über Unterrichtsmaterialien im Mathematikunterricht [Capturing the potential for cognitive activation via teaching materials in mathematics education]. *Zeitschrift Für Erziehungswissenschaft*, 24(4), 955–983. <https://doi.org/10.1007/s11618-021-01020-9>
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Eds.), *Second handbook of research on mathematics teaching and learning. A project of the National Council of Teachers of Mathematics* (pp. 371–404). Information Age.
- Hill, H. C., & Charalambous, C. Y. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Lessons learned and open issues. *Journal of Curriculum Studies*, 44(4), 559–576. <https://doi.org/10.1080/00220272.2012.716978>
- Jentsch, A., Heinrichs, H., Schlesinger, L., Kaiser, G., König, J., & Blömeke, S. (2021a). Multi-group measurement invariance and generalizability analyses for an instructional quality observational instrument. In M. Blikstad-Balas, K. Klette & M. Tengberg (Eds.), *Ways of Analyzing Teaching Quality. Potential and Pitfalls* (pp. 121–139). Scandinavian University Press. <https://doi.org/10.18261/9788215045054-2021-04>
- Jentsch, A., Schlesinger, L., Heinrichs, H., Kaiser, G., König, J., & Blömeke, S. (2021b). Erfassung der fachspezifischen Qualität von Mathematikunterricht: Faktorenstruktur und Zusammenhänge zur professionellen Kompetenz von Mathematiklehrpersonen [Assessment of the subject-specific quality of mathematics teaching: Factor structure and correlations to the professional competence of mathematics teachers]. *Journal Für Mathematikdidaktik*, 42, 97–121. <https://doi.org/10.1007/s13138-020-00168-x>
- Johnson, H. L., Coles, A., & Clarke, D. (2017). Mathematical tasks and the student: Navigating “tensions of intentions” between designers, teachers, and students. *ZDM – Mathematics Education*, 49(6), 813–822. <https://doi.org/10.1007/s11858-017-0894-0>
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., & Kunter, M. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenklassifikation im COACTIV-Projekt* [Classification scheme for mathematics tasks: Documentation of task classification in the COACTIV project]. Max-Planck-Institut für Bildungsforschung.
- Joyce, J., Gitomer, D. H., & Iaconangelo, C. J. (2018). Classroom assignments as measures of teaching quality. *Learning and Instruction*, 54, 48–61. <https://doi.org/10.1016/j.learninstruc.2017.08.001>
- Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M., & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers—Cognitive versus situated approaches. *Educational Studies in Mathematics*, 94(2), 161–182. <https://doi.org/10.1007/s10649-016-9713-8>
- Kieran, C., Doorman, M., & Ohtani, M. (2015). Frameworks and principles for task design. In A. Watson & M. Ohtani (Eds.), *Task Design in Mathematics Education* (pp. 19–81). Springer International Publishing (New ICMI study series). [https://doi.org/10.1007/978-3-319-09629-2\\_2](https://doi.org/10.1007/978-3-319-09629-2_2)
- König, J., Bremerich-Vos, A., Buchholtz, C., Fladung, I., & Glutsch, N. (2020). Pre-service teachers' generic and subject-specific lesson-planning skills: On learning adaptive teaching during initial teacher education. *European Journal of Teacher Education*, 43(2), 131–150. <https://doi.org/10.1080/02619768.2019.1679115>
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621. <https://doi.org/10.1080/01621459.1952.10483441>
- Kullberg, A., Runesson, U., & Mårtensson, P. (2014). Different possibilities to learn from the same task. *PNA: Revista de Investigación en Didáctica de la Matemática*, 8(4), 139–150. <https://doi.org/10.30827/pna.v8i4.6113>
- Kunter, M., & Voss, T. (2013). The model of instructional quality in COACTIV: A multicriteria analysis. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* (pp. 97–124). Springer. [https://doi.org/10.1007/978-1-4614-5149-5\\_6](https://doi.org/10.1007/978-1-4614-5149-5_6)
- Levensen, E., Swisa, R., & Tabach, M. (2018). Evaluating the potential of tasks to occasion mathematical creativity: Definitions and measurements. *Research in Mathematics Education*, 20(3), 273–294. <https://doi.org/10.1080/14794802.2018.1450777>
- Lithner, J. (2017). Principles for designing mathematical tasks that enhance imitative and creative reasoning. *ZDM – Mathematics Education*, 49(6), 937–949. <https://doi.org/10.1007/s11858-017-0867-3>
- Lynch, K., Chin, M., & Blazar, D. (2017). Relationships between observations of elementary mathematics instruction and student achievement: Exploring variability across districts. *American Journal of Education*, 123(4), 615–646. <https://doi.org/10.1086/692662>
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18, 50–60. <https://doi.org/10.1214/aoms/1177730491>

- Matsumura, L. C., Garnier, H., Pascal, J., & Valdés, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment*, 8, 207–229. [https://doi.org/10.1207/s15326977ea0803\\_01](https://doi.org/10.1207/s15326977ea0803_01)
- Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., & Resnick, L. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment (CSE Tech. Rep. No. 681)*. University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide. Seventh Edition*. Muthén & Muthén.
- Neubrand, M., Jordan, A., Krauss, S., Blum, W., & Löwen, K. (2013). Task analysis in COACTIV: Examining the potential for cognitive activation in German mathematics classrooms. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* (pp. 125–144). Springer. [https://doi.org/10.1007/978-1-4614-5149-5\\_7](https://doi.org/10.1007/978-1-4614-5149-5_7)
- Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen: Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie* [A classification of mathematical tasks for analyzing classroom situations: Independent work in student work phases in the lessons of the TIMSS video study]. Franzbecker.
- Niss, M., & Højgaard, T. (2011). *Competencies and mathematical learning – Ideas and inspiration for the development of mathematics teaching and learning in Denmark* (No. 485). Roskilde: MFUFA, Roskilde University. English translation of part I–VI of Niss and Jensen (2002).
- Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, 102(1), 9–28. <https://doi.org/10.1007/s10649-019-09903-9>
- Niss, M. (2015). Mathematical competencies and PISA. In K. Stacey, & R. Turner (Eds.), *Assessing mathematical literacy* (pp. 35–55). Springer. [https://doi.org/10.1007/978-3-319-10121-7\\_2](https://doi.org/10.1007/978-3-319-10121-7_2)
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535–569. <https://doi.org/10.1080/10705510701575396>
- Pettersen, A., & Braeken, J. (2019). Mathematical competency demands of assessment items: A search for empirical evidence. *International Journal of Science and Mathematics Education*, 17(2), 405–425. <https://doi.org/10.1007/s10763-017-9870-y>
- Pettersen, A., & Nortvedt, G. A. (2018). Identifying competency demands in mathematical tasks: Recognising what matters. *International Journal of Science and Mathematics Education*, 16(5), 949–965. <https://doi.org/10.1007/s10763-017-9807-5>
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM – Mathematics Education*, 50(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Prenzel, M., Blum, W., & Klieme, E. (2015). The impact of PISA on mathematics teaching and learning in Germany. In K. Stacey, & R. Turner (Eds.), *Assessing mathematical literacy* (pp. 239–248). Springer. [https://doi.org/10.1007/978-3-319-10121-7\\_12](https://doi.org/10.1007/978-3-319-10121-7_12)
- Ramaswamy, V., DeSarbo, W., Reibstein, D., & Robinson, W. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, 12, 103–124. <https://doi.org/10.1287/mksc.12.1.103>
- Resnick, L. B. (1975). *Task analysis in instructional design: Some cases from mathematics*. National Institute of Education.
- Rezat, S., & Sträßer, R. (2012). From the didactical triangle to the socio-didactical tetrahedron: Artifacts as fundamental constituents of the didactical situation. *ZDM – Mathematics Education*, 44(5), 641–651. <https://doi.org/10.1007/s11858-012-0448-4>
- Ross, N., & Adleff, A.-K. (2022). Anforderungsstufen von Modellierungskomplexität: Transfermöglichkeiten des theoretischen Konstrukts aus dem Aufgabenklassifikationssystem von TEDS-Validierung in Maßnahmen der Lehrkräfteprofessionalisierung [Levels of demand in modeling complexity: Opportunities for transferring the theoretical construct from the task classification system of TEDS-Validate into interventions for teacher professional development]. In N. Buchholtz, B. Schwarz, & K. Vorhölter (Eds.), *Initiationen mathematikdidaktischer Forschung. Festschrift zum 70. Geburtstag von Gabriele Kaiser* (pp. 341–368). Springer. [https://doi.org/10.1007/978-3-658-36766-4\\_18](https://doi.org/10.1007/978-3-658-36766-4_18)
- Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*, 7, 110. <https://doi.org/10.3389/fpsyg.2016.00110>
- Schlesinger, L., Jentsch, A., Kaiser, G., König, J., & Blömeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM – Mathematics Education*, 50(3), 475–490. <https://doi.org/10.1007/s11858-018-0917-5>

- Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189x015002004>
- Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, 33(2), 455–488. <https://doi.org/10.3102/00028312033002455>
- Sullivan, P., & Mornane, A. (2014). Exploring teachers' use of, and students' reactions to, challenging mathematics tasks. *Mathematics Education Research Journal*, 26(2), 193–213. <https://doi.org/10.1007/s13394-013-0089-0>
- Sullivan, P., Clarke, D., & Clarke, B. (2009). Converting mathematics tasks to learning opportunities: An important aspect of knowledge for mathematics teaching. *Mathematics Education Research Journal*, 21(1), 85–105. <https://doi.org/10.1007/bf03217539>
- Thanheiser, E. (2017). Commentary on mathematical tasks and the student: Coherence and connectedness of mathematics, cycles of task design, and context of implementation. *ZDM – Mathematics Education*, 49(6), 965–969. <https://doi.org/10.1007/s11858-017-0895-z>
- Turner, R., Dossey, J., Blum, W., & Niss, M. (2013). Using mathematical competencies to predict item difficulty in PISA: A MEG study. In M. Prenzel, M. Kobarg, K. Schöps, & S. Rönnebeck (Eds.), *Research on PISA* (pp. 23–37). Springer. [https://doi.org/10.1007/978-94-007-4458-5\\_2](https://doi.org/10.1007/978-94-007-4458-5_2)
- Turner, R., Blum, W., & Niss, M. (2015). Using competencies to explain mathematical item demand: A work in progress. In K. Stacey, & R. Turner (Eds.), *Assessing mathematical literacy. The PISA experience* (pp. 85–115). Springer: Cham. [https://doi.org/10.1007/978-3-319-10121-7\\_4](https://doi.org/10.1007/978-3-319-10121-7_4)
- Turner, R., Tout, D., & Spithill, J. (2023). A rich view of mathematics education and assessment: Mathematical competencies. In U. T. Jankvist, & E. Geraniou (Eds.), *Mathematical Competencies in the Digital Era* (pp. 329–354). Springer. [https://doi.org/10.1007/978-3-031-10141-0\\_18](https://doi.org/10.1007/978-3-031-10141-0_18)
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars, & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge University Press. <https://doi.org/10.1017/cbo9780511499531.004>
- Watson, A., & Ohtani, M. (2015). Themes and issues in mathematics education concerning task design: Editorial introduction. In A. Watson, & M. Ohtani (Eds.), *Task Design In Mathematics Education* (pp. 3–15). Springer International Publishing (New ICMI study series). [https://doi.org/10.1007/978-3-319-09629-2\\_1](https://doi.org/10.1007/978-3-319-09629-2_1)
- Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class analysis: A guide to best practice. *Journal of Black Psychology*, 46(4), 287–311. <https://doi.org/10.1177/0095798420930932>
- Yaro, K., Amoah, E., & Wagner, D. (2020). Situated perspectives on creating mathematics tasks for peace and sustainability. *Canadian Journal of Science, Mathematics and Technology Education*, 20, 218–229. <https://doi.org/10.1007/s42330-020-00083-w>

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

## Authors and Affiliations

Ann-Kristin Adleff<sup>1</sup> · Natalie Ross<sup>1</sup> · Johannes König<sup>2</sup>  · Gabriele Kaiser<sup>1,3</sup> 

✉ Gabriele Kaiser  
gabriele.kaiser@uni-hamburg.de

Ann-Kristin Adleff  
ann-kristin.adleff@uni-hamburg.de

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

<sup>2</sup> Faculty of Human Sciences, University of Cologne, Gronewaldstraße 2, 50931 Cologne, Germany

<sup>3</sup> Faculty of Education and Arts, Nord University, Universitetsalléen 11, 8026 Bodø, Norway