ZDM – Mathematics Education (2024) 56:981–996
https://doi.org/10.1007/s11858-024-01626-3

ORIGINAL PAPER

Mathematics teachers' multiple perspectives on adaptive tasks: task evaluation and selection as core practices for teaching quality

Thomas Bardy[1,2] · Lars Holzäpfel[1] · Frank Reinhold[1] · Timo Leuders[1]

Accepted: 10 August 2024 / Published online: 28 August 2024

Abstract

The selection of tasks based on the evaluation of task features can be considered a core practice of teaching and a relevant component of teaching quality. This is typically part of teachers' preparation for their classroom teaching, which prompts the following question: What are the characteristics of the tasks that teachers use when selecting tasks for differentiated teaching? To answer this question, we analyzed systematic differences in the focus of 78 in-service high school and lower secondary school teachers during the evaluation of task features. The teachers had to select eight tasks about the practice of fractions with respect to their differentiation potential—operationalizing their adaptive teaching competence from a mathematics educational perspective. To analyze the differences, we performed a cluster analysis of the task features that the teachers drew upon. Three groups of teachers could be identified with variations in their focus on directly or indirectly relevant, domain-specific or domain-general task features. Taking into account such variations may explain differences in teaching quality and student outcomes and may be relevant when designing teacher professional development programs.

Keywords Differentiation potential · Classroom heterogeneity · Adaptive tasks · Task features · Hierarchical cluster analysis · Teacher professional development

1 Introduction

Tasks play a central role as learning opportunities in mathematics classrooms (Brod, 2024; Bromme, 1981; Chapman, 2013; Hiebert et al., 2003a; Johnson et al., 2017; Krauss et al., 2008; Thanheiser, 2017). They specify the topic of a learning situation and strongly contribute to the subsequent learning processes (Hiebert et al., 2003b; Stein et al., 1996). Thus, the selection of tasks based on the evaluation of task features can be considered a *core practice* of teaching and a relevant component of teaching quality (Forzani, 2014; Grossman et al., 2009).

The selection of tasks is generally part of teachers' preparation for their classroom teaching and is guided by reasoning about specific task features—which is the focus of the present article. Of course, the selection of tasks is only a first step, which may be followed by processes of task adaptation, the design of an instructional setting, the enactment of the task in class discussions and the evaluation of students' solution attempts (Brod, 2024; Thanheiser, 2017).

Within this complex context of tasks in the mathematics classroom, there has been much research on task construction in general (Sullivan et al., 2013) and specifically on task implementation by teachers in the classroom (Hiebert et al., 2003b; Stein et al., 1996); however, much less attention has been allocated to the selection of tasks by teachers during classroom preparation. Although there is ample evidence of the deficiencies of the tasks selected (e.g., for Germany, Adleff et al., 2023; Jordan et al., 2008) and some research on teachers' reasoning when evaluating given tasks (Hammer & Ufer, 2023; Philipp, 2018; Pilous et al., 2023), we still have limited knowledge of the process of task selection when several tasks are assigned and also of the teachers' reasoning in doing so (e.g., Rieu et al., 2022). The selection process is of interest for practical reasons: In many school systems, teachers are rather free to select from different materials, and for theoretical and methodological reasons, teachers' judgment and decision

✉ Thomas Bardy
thomas.bardy@fhnw.ch

[1] University of Education, Freiburg, Germany

[2] University of Applied Sciences and Arts Northwestern Switzerland, Windisch, Switzerland

making can be investigated by systematically varying the tasks and their features used in an experimental design.

For reasons of feasibility, we focus on such task features that are relevant for dealing with heterogeneity in classes. There is a special requirement for tasks to have a high *differentiation potential* (we will also refer to such tasks as *adaptive tasks*) to achieve adaptivity to different student levels (Gallagher et al., 2022). However, only a few tasks in mathematics textbooks demonstrate this differentiation potential (Bardy et al., 2021; Heinle et al., 2022), and teachers do not easily identify and use adequate tasks (Hammer & Ufer, 2023; Leuders & Föckler, 2016); both phenomena strongly influence the quality of mathematics instruction from a content-specific perspective.

Therefore, the present study investigates teachers' appropriate selection of tasks based on specific features as an operationalization of a core practice. Teachers were asked to articulate which features of tasks they use when judging and selecting adaptive tasks (Fig. 1). We elaborate on how some of these task features are *domain-specific* or *domain-general*, thus constituting two complementary views regarding teaching quality (Dreher & Leuders, 2021; Praetorius & Charalambous, 2018). We present a theoretical framework for categorizing task features with respect to the differentiation potential of tasks from this perspective, and we investigate whether teachers' focus varies systematically in a way that allows us to identify groups of teachers based on differences in the quality of situated task selection.

Our study contributes to addressing how teachers' perspectives regarding teaching quality vary with respect to domain-specific and domain-general features, shedding light on task evaluation and selection as core practices to explain differences in the quality of the preparation of mathematics instruction.

1.1 Task features regarding the differentiation potential of tasks in mathematics education research

When appropriately selected (and implemented), mathematical tasks are essential learning opportunities for students (Johnson et al., 2017; Krauss et al., 2008; Thanheiser, 2017). Mathematics education research on tasks frequently focuses on the presence or absence of specific features that tasks can encompass (Jordan et al., 2006; Yeo, 2017). Such features may be based on domain-specific or domain-general theories regarding their perceived benefits (or hindrances) for learning. The high degree of *openness* of tasks is believed to allow for more diverse problem-solving strategies (Anderson, 1996; Sullivan, 1999; Yeo, 2017); the differing *cognitive requirements* of tasks may explain the systematic difficulties of students (Hammer, 2016; Maier et al., 2010; Stein & Smith, 1998; Stein et al., 1996); a high *motivational potential* of tasks can be related to students' engagement (Heinle et al., 2022); tasks may or may not be suitable to foster specific students' *competencies* (Bölsterli Bardy & Wilhelm, 2018); tasks may also exhibit features that relate

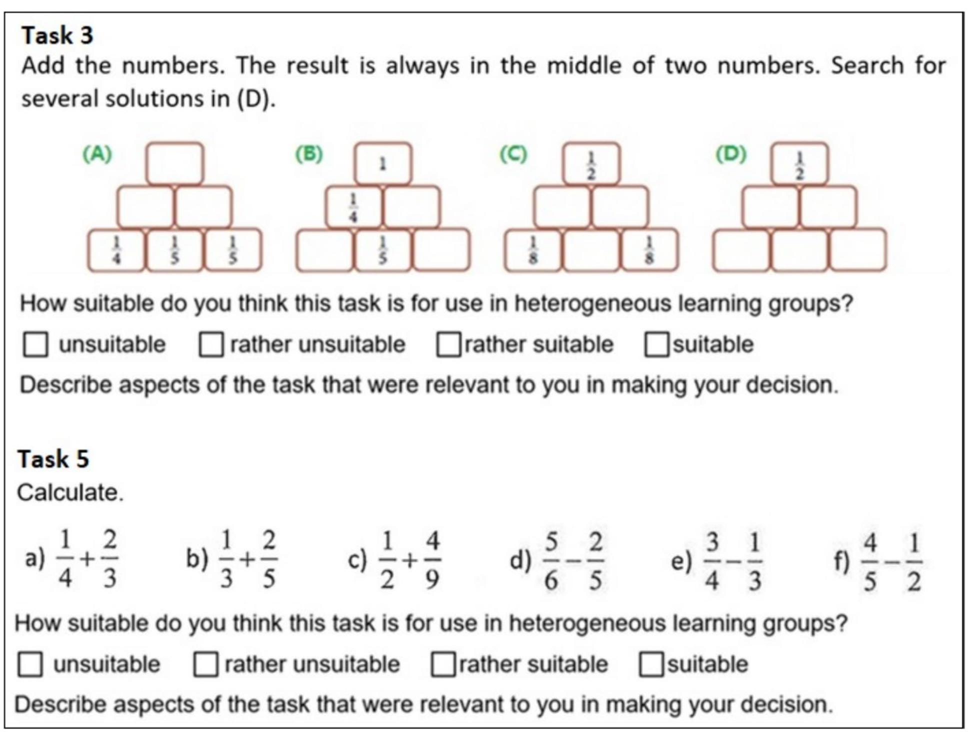

Fig. 1 Examples of tasks 3 and 5 from the questionnaire used in the present study, illustrating task features regarding the differentiation potential of tasks

to how effectively they support and promote the *learning processes* of students (Blömeke et al., 2006).

Of particular relevance for the present study are task features regarding the inherent differentiation potential of tasks, as in whether a task can be worked on at different depths and at different cognitive levels and whether students themselves—not teachers—can decide how to proceed (Blömeke et al., 2006; see also Leuders & Prediger, 2016; Wellenreuther, 2004).

For instance, if a practice phase is implemented after the introduction of fraction addition, the key questions for teachers regarding the differentiation potential of the two tasks presented in Fig. 1 may be as follows: Can a generally low-achieving student practice with the task? Can a generally high-achieving student also find opportunities to practice with the task? Can both students work on the same task at the same time?

One means to find informed answers to these questions is to evaluate task features (Anderson, 2003; Bardy et al., 2021; Hammer & Ufer, 2023; Yeo, 2017). The two given tasks (total of 8 tasks) could be distinguished on the basis of their a*ccessibility* (Stillman, 2004), which indicates whether a task contains introductory parts that enable low-achieving students to start successfully, such as understanding the task, simple cases, and concrete examples (Fig. 1, task 3, provides a simple subtask (A) that only requires addition and working forward; a task in which a worked example is presented would be considered to be more accessible). Subtask (D) in task 3 (Fig. 1) allows the search for multiple solutions, which can be performed on a simple and more elaborate level—exemplifying feature *openness* (Yeo, 2017). In contrast, task 5 (Fig. 1) lacks these features. Further features that contribute to the differentiation potential of a task may be elements of *guidance* (such as the hints at the beginning of task 3; Lazonder & Harmsen, 2016) or *support* simplifying certain solution steps. Although such features could be added by a teacher during the lesson, they are useful when they are already part of the task text when they work individually in "seatwork" (Neubrand, 2006).

In addition to the aforementioned features regarding the differentiation potential of tasks present in the literature, our own previous research endeavors aimed to empirically extract task features that teachers actually mention when being asked to judge the differentiation potential of several tasks on the topic of fractions (Bardy et al., 2021), including the following: Task feature *difficulty* refers to the activation of cognitive processes during task processing and is of central importance for identifying the differentiation potential of a task; tasks that allow for working on different levels of difficulty are potentially more adequate for a heterogeneous group of students. Task feature *openness* describes the extent to which students can make decisions regarding a goal of a task (i.e., *what* would be considered an appropriate solution for the task) or a problem-solving strategy (i.e., *how* to reach an appropriate solution). *Language* is also an important feature of the differentiation potential of a task (Leuders & Prediger, 2016; Wessel & Erath, 2018). A differentiating linguistic formulation in a task can support low- and high-achieving students in understanding the task. Other features, such as *support* and *guidance,* can be realized in a task, particularly to help low-achieving students complete the task (Blömeke et al., 2006; Jordan et al., 2006; Wellenreuther, 2004).

Bardy et al.'s (2021) qualitative analysis has successfully identified a diverse range of task features regarding differentiation potential, which is actually accounted for by teachers during the process of task selection. However, this previous approach primarily serves to categorize these features but does not indicate their interrelation during task selection, such as their relative importance for teachers or even different groups of teachers. Consequently, while we gained insight into the various features perceived as significant, our previous study did not explore the interrelationships among these characteristics or yield insights into which constellation of features in tasks teachers focus on—leaving the question regarding qualitative differences in their teaching quality unanswered.

1.2 Evaluation of task features and task selection

Selecting, developing, and implementing mathematical tasks are distinct but related facets of pedagogical content knowledge (Ball et al., 2001). The selection of tasks based on an evaluation of task features as a core practice of teachers can occur during the preparation of teaching (e.g., with the aim of selecting appropriate tasks) or while reflecting on the teaching (e.g., by evaluating whether the task adequately contributes to the teaching goals). The process of evaluating task features is also described as *task assessment*, *task analysis*, or *task diagnosis* (Bromme, 1981; Hammer & Ufer, 2023; Smith & Stein, 1998).

1.2.1 Teachers' cognitive processes

The extent to which teachers are guided by such features related to the *deep structure* of tasks has been investigated in several studies (Anderson, 2003; Bardy et al., 2021; Hammer & Ufer, 2023), suggesting that there is no common view on adaptive tasks among teachers. However, the evaluation of task features is considered to be a knowledge-based process (Anderson, 2003; Hammer & Ufer, 2023; Seidel et al., 2010; van Es & Sherin, 2002). Regarding the quality of mathematics instruction, a relevant question concerns how teachers make use of their knowledge when evaluating and selecting tasks during the preparation of their classroom lessons. Quantitative studies assessing the influence of knowledge components (i.e., content knowledge and pedagogical content knowledge) on the evaluation of task features have yielded diverse results.

Hammer (2016) empirically distinguished between two distinct knowledge-based processes, namely, the *identification* of the task potential and the *analysis* of the task potential; the author focused on selection without a specific focus on the differentiation potential of tasks. Ostermann et al. (2018) reported that (a) specific knowledge of difficulty-generating task features allows teachers to identify relevant task features, resulting in more accurate judgments of absolute and relative task difficulty, whereas (b) unspecific knowledge regarding the general difference between their expert knowledge and students' knowledge can help teachers' anchoring in reducing the common underestimation of task difficulty. Rieu et al. (2022) assume that during task selection, teachers identify task features that are relevant (or irrelevant) to a task's difficulty and then integrate information regarding multiple features based on their knowledge of the relevance of these features. These assumptions were supported empirically by testing predictions concerning the influence of knowledge and time pressure. Pilous et al. (2023) reported that teachers, to varying degrees and depending on their professional backgrounds (i.e., experts in teaching mathematics, in teaching mathematics education, or in mathematics), either refer more to general pedagogical aspects or to domain-specific aspects (i.e., from the domain of mathematics) and topic-specific aspects (i.e., regarding specific mathematical topics such as numbers, geometry, functions, etc.) relevant to task selection and implementation.

This research on (latent) cognitive processes or (manifest) arguments of teachers during the evaluation of task features amounts to the assumption that the *core processes* of task selection are (1) identifying features, (2) relating these features to a specific goal (in the present study: the differentiation potential of tasks), and (3) integrating these judgments into an overall judgment of the adequacy of a task and/or a decision about selecting or modifying a task. The present study focuses on the first two aspects. Therefore, the presented findings indicate that knowledge of task features supports an accurate estimation of task difficulty. They further suggest that the decision process indeed requires the recognition and combination of several features in a task. However, the present quantitative studies focused on highly content-specific task features, such as the possibility of certain misconceptions with respect to functions and graphs, and they did not specifically address the differentiation potential.

1.2.2 Differences between teachers during the evaluating of tasks

This study focuses on task evaluation during teaching preparations. For that purpose, teachers anticipate students' thinking in terms of emerging comprehension difficulties, specific requirements, or the solution process, for instance (Morris et al., 2009; Philipp, 2018). Regarding the evaluation of a task's differentiation potential, teachers may recognize that a task already has a high degree of *difficulty* at the beginning of the solution process (that their low-achieving students may not be able to cope with) or that the *openness* of a task offers the potential that their low-achieving students will find a solution that corresponds to their respective performance level; they may also remember misconceptions their students held when working on similar tasks.

Teachers may consider various features of tasks for their individual evaluation, which may result in different task selections. One question that research on mathematics teacher education has attempted to answer concerns whether there are coherent groups of teachers that focus on comparable subgroups of task features—leading to specific patterns in task selection. Hammer (2016) identified three groups of teachers: The "learning process-oriented group" is characterized by a strong orientation toward cognitive processes and the interconnectedness of mathematical ideas. The "learner-oriented group" cites task features in their rationale for task selection, such as increasing motivation through the task context, allowing for differentiation with the task, or enabling a sense of accomplishment. The "practice-oriented group" focuses on their justifications for mathematical activities (e.g., practicing or securing results), on organizational aspects related to the task, and on the clarity and structure of the task (Hammer, 2016, p. 159). In all three groups, content-specific and domain-general aspects are closely connected. Hammer also stresses that teachers differ in terms of whether they refer to the surface and deep structure of tasks. Teachers who had participated in a specific training regarding the handling of tasks described tasks primarily with features of the *deep structure* (Hammer, 2016, p. 185); teachers who had not participated in the training mentioned features of the *surface structure* of the tasks (including practicing, organizational aspects, and generic task features). Kuntze (2011) demonstrated that prospective teachers assigned lower learning potential to tasks with increased modeling demands relative to tasks in which the mathematical model was already largely determined in advance. Compared with in-service teachers, these teachers clearly rated the learning potential of the tasks presented more similarly; however, there was no advantage of tasks with greater modeling demands. Dreher (2015) investigated teachers' ability to recognize when "tasks support the learning process" in fraction tasks via either the conversion of representations or unhelpful pictorial representations. Recognizing this focus is relevant for differentiation decisions. A cluster analysis was used to identify different teachers' views regarding the learning potential of tasks.

1.3 The present study

The selected task and the verbal justification referring to task features provide us with information about the assumed task selection process. On the one hand, a teacher's product can deviate from expert judgment. In this case, the task features explicitly mentioned convey information about how the teachers deviated from the expert judgment in the task selection process regarding the recognition of the differentiation potential. Alternatively, if the justification concurs with the expert judgment, verbal justification may still be evident regarding whether teachers addressed relevant task features in the task selection process. This evaluation may still suffer from misinterpretations when the teachers do not reliably verbalize their reasoning.

The question regarding which task features teachers consider relevant when evaluating the differentiation potential of tasks has only partly been answered. Some features may be more important to teachers than others are, although they may be less relevant for differentiation. This raises the question of whether teachers apply an accurate focus toward relevant task features. There is no simple hierarchy of competencies in this respect. Rather, teachers differ in their profiles when they recognize and judge task features, indicating that a person-centered approach is needed to investigate teachers' choices. Are there groups of teachers who differ systematically when evaluating multiple dimensions of task features? These questions guided our research in the present study.

1.3.1 A systematic characterization of task features related to differentiation potential

For that purpose, a systematic characterization of task features with respect to differentiation potential is necessary. Regarding two complementary views of teaching quality (Dreher & Leuders, 2021; Praetorius & Charalambous, 2018), we argue that task features can be rooted in a *domain-specific* analysis of a task, or an analysis utilizing mathematics education-related categories and knowledge facets to describe the task (e.g., task feature openness, goal differentiation, language, or motivation), which we distinguish from a *domain-general* perspective on the differentiation potential of tasks (e.g., feature guidance, support, layout, or presentation). *Directly relevant* features are essential to the differentiation potential of tasks, inherent to the tasks themselves, and persistently present regardless of how they are implemented in the mathematics classroom (e.g., the feature's openness, goal differentiation, guidance, or support), whereas features that are *indirectly relevant* for the differentiation potential of tasks can contribute to adaptivity but do not necessarily have to; the potential of the latter-mentioned features is dependent on the implementation in the classroom and/or the connection to other tasks and therefore not bound to the task (e.g., the feature's language, motivation, layout, or presentation).

Table 1 Domain-specific vs. domain-general task features and task features with direct vs. indirect relevance for the differentiation potential of tasks

Task features	Domain-specific	Domain-general
Directly relevant	**Category *Specific adaptivity***	**Category *Task structure (1)***
	Openness	Guidance
	Accessibility	Support
	Goal differentiation	Self-checking
	Difficulty	
Indirectly relevant	**Category *Task content***	**Category *Task structure (2)***
	Routine/drill	Size
	Creativity	Layout
	Reasoning	Presentation
	Problem solving	Transparent structure
	Generalizing	
	Context/application	
	Language	
	Motivation	
	Prerequisites	
	Change of representation	

All the task features considered in the present study are presented in Table 1 (see the appendix for a detailed description of each feature).

Thus, in the *domain-specific* and *directly relevant* cells, we collect the features that are primarily important for the differential potential of tasks; among them, *openness* (as described above), *accessibility* (a task containing introductory parts that facilitate successful entry, particularly for low-achieving students), *goal differentiation* (the level up to which the goal state of a problem can be freely selected among several potentially relevant goal states), and *difficulty* (as described above) can be identified. These features of category *specific adaptivity* (AS) are closely connected to the mathematical characteristics of the item; thus, they are considered *domain-specific*, and they affect the differentiation potential directly, and independently of implementation in the classroom.

We consider task features *domain-specific* yet *indirectly relevant* for the differentiation potential of tasks when they refer to the content of the task. Whether the task addresses *routine or drill* questions and stimulates *creativity*, *reasoning*, or *problem-solving* activities does not directly influence whether a task is inherently adaptive—the differentiation potential of such features depends strongly on the implementation of the teacher in the classroom and has no inherent task characteristics. In addition, the *context* of the real-world application in which the task is situated, the *level of language* (as described above), or further content-related features, such as *changes in mathematical representations* between or within the item, also account for mathematical aspects of the tasks but not primarily their differentiation potential. We refer to the second domain-specific category of task features from the deep structure as *task content* (TC).

Domain-general and *directly relevant* features are *self-regulating* features, such as *guidance* (a task containing concrete instructions about how to proceed in finding a solution), *support* (a task containing additional help, such as an explanation, a concrete example, or a visualization), and *self-checking* (a task already indicating the solution or intermediate results for later control)—they inherently address adaptivity to students' needs independently of task implementation. Furthermore, we consider the *domain-general* surface structure of the task (i.e., its *layout* or its *structure*) to be only *indirectly relevant* to the analysis of the differentiation potential. For the present study, we refer to both categories of domain-general task features as *task structure* (TS).

In conclusion, the differentiation potential of a task is directly or indirectly influenced by exceptionally different domain-specific or domain-general task features. It already becomes apparent that many of the content-specific features refer to aspects that are not discernible on the surface structure of a task but require reflection upon an imagined solution process and thus can be considered features of the deep structure of a task—which indicates a higher level of expertise among such teachers, who would primarily use such features.

1.3.2 Research questions

To the best of our knowledge no empirical studies exist currently that relate the domain-specific dimension to other dimensions of task feature evaluation. However, this dichotomy is especially important when the domain-general aspects of the quality of mathematics instruction are considered. We expected that teachers would differ systematically with respect to focusing on the surface or deep structure of tasks—and thus on different domain-specific or domain-general task features. This extends previous research by identifying systematic differences in teachers' task evaluation and selection—situated around the practically relevant context of a task's differentiation potential.

RQ1 *Does the focus on different task features vary systematically across different subgroups of teachers?* This question will be answered via a cluster analysis based on data from in-service mathematics teachers.

RQ2 *What are the characteristics of the reasoning processes of prototype teachers from identified subgroups in relation to specific task features?* We employ qualitative descriptions of actual teachers from the sample to verify the cluster centers obtained through exploratory quantitative clustering methods, focusing on the criteria used for clustering.

RQ3 *To what extent can teachers with varying emphases on task characteristics provide an accurate assessment of the overall suitability of a task for adaptive exercise?* We compare the judgments of teachers in each of the resulting clusters to an expert norm. Our hypothesis is that clusters with a stronger focus on domain-specific and directly relevant task features are more accurate in assessing the task's differentiation potential.

RQ4 *Does the general profile of teachers (e.g., gender, teaching experience, employment level, or school type) correlate with their focus on distinct task-related features?* We expect systematic differences in the relationships of personal and job-related characteristics to the clusters; for instance, teachers from lower school tracks may face a greater need for adaptive mathematics instruction in their daily work so that they may have a broader background in relation to adaptive tasks compared to teachers from higher school tracks. In particular, the variability of teachers from different types of schools in Germany could precipitate different focuses on task characteristics, as teachers in lower secondary schools in Germany are more closely related to adaptive tasks (e.g., through their textbooks) than teachers in other types of schools.

2 Methods

2.1 Sample

The participants in our study were $N=78$ in-service mathematics teachers (46 female and 32 male) in Germany; the German teacher education system differs between high school teachers (who have a more in-depth education in mathematics) and lower secondary school teachers; 33 teachers teach at a lower secondary school, and 45 teachers teach at a high school, with different lengths of teaching experience (3 teachers with less than or equal to 3 years; 12 teachers with 3 to 9 years; 24 teachers with more than 9 years, 39 teachers who did not report) and different levels of employment (3 teachers with less than 50%; 3 teachers with exactly 50%; 9 teachers with 50% to 100%; 26 teachers with exactly 100%; 37 teachers who did not report). The teachers participated voluntarily and had all registered for a teacher-training course on the topic of "differentiation with tasks in mathematics lessons," which lasted for one year; all the data were taken from an assessment before the very first training session.

2.2 Instrument and procedure

We used exercise tasks that had to be evaluated in terms of their suitability for use in heterogeneous learning groups during exercise phases (e.g., Figs. 1, 2). The assessment employed a paper–pencil questionnaire comprising eight tasks, each with varying potential for differentiation.

2.2.1 Judging the differentiation potential of tasks

For each task, an overall judgment of the task's differentiation potential was assessed with the following question: "How suitable do you think this task is for use in heterogeneous learning groups?" (Fig. 2) A four-point Likert scale was used. This judgment was compared to an expert norm: Three of the authors (former mathematics schoolteachers with experience at the classroom level and as researchers in mathematics education working on the topic of differentiation for several years) individually judged the tasks with good agreement (Fleiss $\kappa=0.54$; Bardy et al., 2021).

2.2.2 Reasoning about task features

For each task, we assessed the reasoning behind the decision-making process by presenting the following prompt via an open-answer format: "Describe aspects of the task that were relevant to you in making your decision" (Fig. 2). Answers were categorized regarding whether they identified domain-specific (vs. domain-general) task features and task features with direct (vs. indirect) relevance for the task's differentiation potential (Table 1). For the full coding manual, refer to the appendix. The intercoder reliability of six raters (i.e., mathematics students) yielded substantial agreement (Fleiss $\kappa=0.65$).

We illustrate which potential views of the example task presented in Fig. 2 can be expected. In such parallel adaptive tasks, two different exercise levels are clearly delineated by a vertical dashed line. One is designed for low-achieving students (on the left), and the other is designed for high-achieving students (on the right). On both sides of the task,

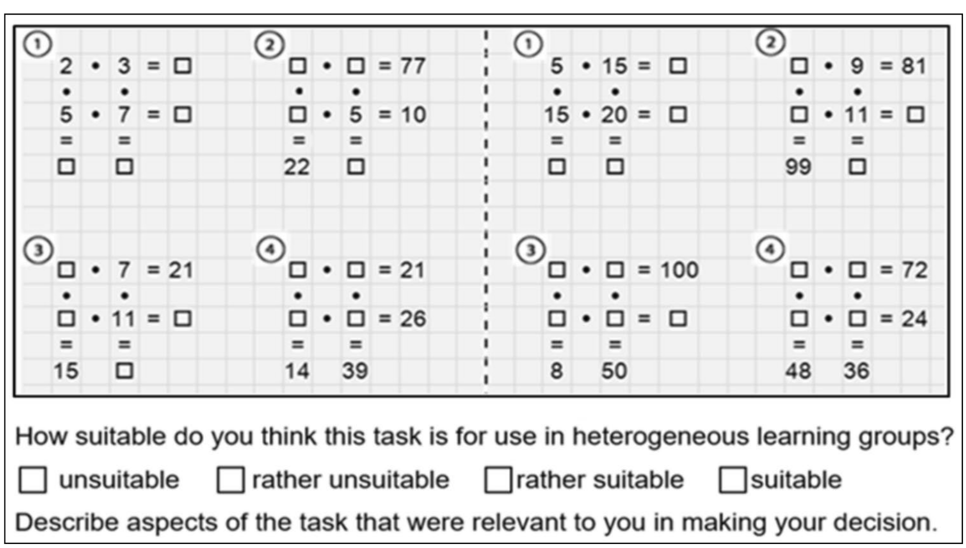

Fig. 2 An adaptive task structured as a parallel adaptive task for low-achieving (left-hand side) and high-achieving (right-hand side) students (Leuders & Prediger, 2016, p. 121)

feature *accessibility* (domain-specific and directly relevant) is notable: Subtask (1) is easy to solve (because it involves only single-digit multiplication in one direction) and serves as an entry point in the exercise phase. Additionally, the task feature *self-checking* (domain-general and directly relevant) is identifiable: Students can identify inconsistencies or errors in their calculations as they progress through the exercise; for instance, if a student fills two gaps in the problem and encounters an issue in the third, this discrepancy prompts the student to retrace their first calculations. Furthermore, on the left-hand side, subtasks (1) to (4) increase in difficulty (domain-specific and directly relevant); for instance, subtask (1) involves straightforward multiplication, whereas subtasks (2) and (3) require students to perform operations in reverse, and subtask (4) provides results without intermediary values—necessitating a complete backtracking of calculations. On the right-hand side, subtasks (1) to (3) are more difficult than subtasks on the left-hand side. In this example, there is coherence between the deep structure and the surface structure of the task: The dashed line, as an element of the task's *layout* (domain-general and indirectly relevant), separates the easier subtasks from the harder subtasks. Additionally, teachers can recognize that subtask (4) on the right-hand side (in contrast to subtask (4) on the left-hand side) is *open* (domain-specific and directly relevant); it allows for multiple solutions.

Based on the collected data, we were able to determine how often a teacher mentioned one of the 21 task features (Table 1). To analyze systematic differences between the teachers on a quantitative level, we combined the two *domain-general* categories (Table 1) so that *each* category could be found by the teachers in *each* task. The resulting category system differentiates between *specific adaptivity* (AS; domain-specific and directly relevant task features), *task content* (TC; domain-specific and indirectly relevant), and *task structure* (TS; domain-general and directly or indirectly relevant; see Sect. 1.3.1).

2.3 Data preparation and analysis

After rating the teacher comments, we calculated the sum scores of the ratings of the categories (AS, TC, TS). Since the various categories were plausibly applicable through tasks with different frequencies, we normalized the teachers' numbers of identifications of each category so that the values entering the cluster analysis all had a range between 0 and 1 (Table 2).

Analyses were conducted in R. We clustered teachers utilizing standardized values of the three categories and a two-step clustering approach with the package {NbClust} (Charrad et al., 2014). First, hierarchical clustering was performed with the Ward method and Euclidean distance measure to identify the appropriate number of clusters according to the majority rule; approximately 30 different stopping rules were used to determine the appropriate number of clusters, as the one with the most stopping rules resulted. Clusters were then defined with the *k*-means algorithm (RQ1). Each teacher was assigned to a specific cluster in terms of his or her references (proportions) to the three categories. Within a cluster, similar profiles can thus be identified (RQ2). To assess the overall judgment of the eight tasks in terms of their differentiation potential, the difference between the mean expert judgment of a task and each participant's judgment is calculated. The mean scores per participant are then estimated, and one-sample t tests against 0 are utilized (RQ3). The relationships between categorical covariates and clusters are evaluated through a comparison of the relative frequency of each covariate's level within each cluster via the chi-square test (RQ4).

Table 2 Overall mean scores on and correlations between the three groups of task features

Category			Correlation		
	M	*SD*	AS	TC	TS
Specific adaptivity (AS)	0.55	0.28	–		
Task content (TC)	0.54	0.22	−0.26*	–	
Task structure (TS)	0.50	0.30	0.10	−0.11	–

*Significant with $p < 0.05$ (two-tailed)

3 Results

3.1 Subgroups of teachers focusing on specific task features

Our first question concerned whether subgroups of teachers exist that differ systematically in their focus on task features (RQ1). We performed a *k-means cluster analysis* to find an appropriate cluster solution. To determine the optimum number of clusters, the hierarchical Ward-Method was first used here. Six of the 23 indices suggest 2 as the ideal number of clusters, while five indices point to 3, and fewer indices point to other numbers of clusters. Consequently, we examined both the two-cluster and three-cluster solutions, as they were the most strongly supported by the majority of indices. The two-cluster solution would have merged two of the three clusters that have distinct characteristics and implications for research on mathematics teachers. Therefore, we opted for the three-cluster solution due to its superior interpretability and relevance (Fig. 3).

Teachers assigned to *Cluster 1* had a *broad focus on all three types of task characteristics*; they presented elevated scores in the categories AS (domain-specific and directly relevant) and TS (domain-general and directly or indirectly

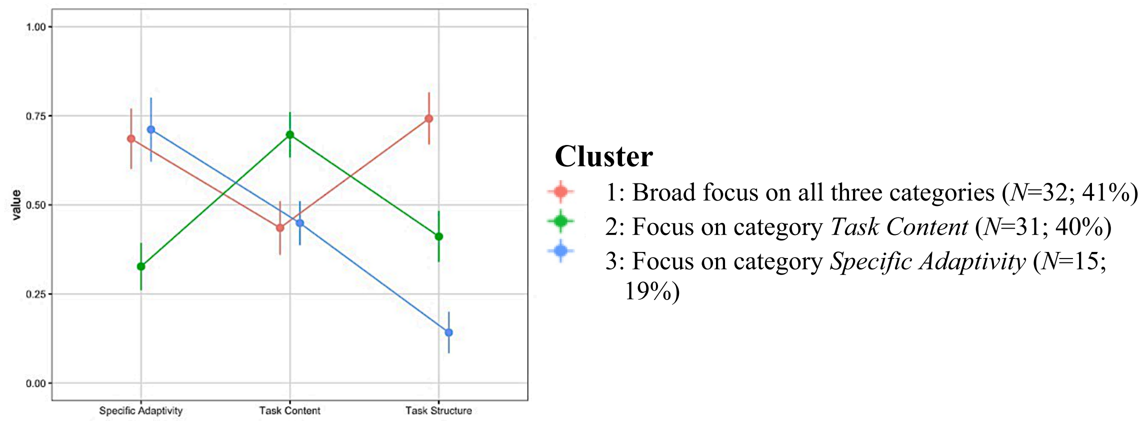

Fig. 3 The three-cluster solution

relevant). Teachers who presented elevated scores in *all* three categories can also be found in Cluster 1. As illustrated in Fig. 3, nearly half of the teachers (41%) were in Cluster 1. *Cluster 2* (40%) consisted of teachers who focused on TC (domain-specific and indirectly relevant) but had low scores in the other two categories. Teachers with elevated scores in the category AS but low scores in the category TS are grouped in *Cluster 3;* they (19%) focused exclusively on domain-specific and directly relevant task features, making it an important cluster for our study.

3.2 Reasoning processes of prototype teachers from these subgroups

To gain deeper insight into the three clusters identified, we will now qualitatively describe three teachers (one from each cluster) in terms of their prototypical responses (RQ2). In general, teachers in Cluster 1 strongly argue with a broad spectrum of task features when analyzing tasks regarding differentiation potential; features such as openness, accessibility, language, difficulty, and layout received equal focus. Megan (AS = 0.9; TC = 0.73; TS = 0.75) is regarded as a prototypical teacher from Cluster 1 (Table 3). Teachers in Cluster 2 described task features that relate to the knowledge required in the task (e.g., routine/drill, creativity, reasoning)—Jack (AS = 0.18; TC = 0.73; TS = 0)—and were prototypical teachers from this cluster (Table 3). Teachers in Cluster 3 almost exclusively referred to task features that explicitly refer to the differentiation potential (e.g., accessibility, openness, difficulty). We considered Betty (AS = 0.99; TC = 0.53; TS = 0) to be a specific representative.

Consider task 0, which is presented in Fig. 2. All three selected prototypical teachers exhibited different reasoning to determine whether this task was suitable for use in heterogeneous classrooms, and their mentioned task features reflected the cluster to which they were assigned (Table 3). For a more detailed qualitative description of the underlying reasoning processes, we have compared Betty (Cluster 3, strong focus on AS) to a fourth prototypical teacher, Noah (Cluster 1, AS = 0.45; TC = 0.27; TS = 0.875; broad focus on all categories of task features).

Noah concluded that task 0 has a low differentiation potential; he justified his evaluation with features from TS: "Not recognizable according to which order the tasks are to be processed. The results feels too small (font size!). Equal signs in the confusing direction. (sic)." From the expert's point of view, this category does not necessarily target the relevant task features. This rather strong focus on the surface structure of the task cannot be expected to capture the relevant features of the task—leading to an evaluation that is not in line with the expert rating of task 0 being suitable for use in heterogeneous classrooms. In contrast, Betty justified her assessment with features from the category AS (Table 3), which we consider to be the adequate features to focus on when adequately evaluating the potential of this task—a characteristic of Cluster 3. However, she still concluded that task 0 is not suited for a differentiating use in class—contrary to the experts.

This case indicates that a focus on the relevant features of a task alone does not necessarily lead to an appropriate overall evaluation of a task—which stresses questions regarding whether a focus on relevant task features generally correlates with accurately assessing the overall suitability of a task for adaptive exercise.

3.3 Performance of the clusters regarding overall task evaluation

As mentioned, we asked to what extent teachers focusing on different task features (i.e., the cluster) exhibit systematic differences in their accuracy in assessing the overall differentiation potential of tasks (RQ3). Table 4 indicates that teachers who are assigned to Cluster 3 generally also rated the tasks more in line with the expert rating compared to teachers from other clusters; they displayed the least deviation from the expert rating scores, and there were (marginal) significant differences between the expert rating and the ratings of Clusters 1 and 2 (Table 4). Table 5

Table 3 Focused task features of Megan, Jack, and Betty in the analysis of task 0 (see Fig. 2)

Teacher	Comparison of the rating of task 0			To be found in…
Megan	**Decision:** Task 0 is "rather unsuitable"	**Feature**	**Category**	
Reasoning	Results are unambiguous; hardly any different calculation methods are possible	Openness	AS	**Cluster 1**
	Access to the task is the same for all students	Openness	AS	
	"Only" difficult numerical values	Difficulty	AS	
	"Only" reproduction of multiplication and division	Routine/Drill	TC	
	Subtask (4) enables deeper understanding	Prerequisites	TC	
	The "task structure" can confuse students	Presentation	TS	
Jack	**Decision:** Task 0 is "rather suitable"			
Reasoning	Different requirement levels	Openness	AS	**Cluster 2**
	Contains "inverse task"	Difficulty	AS	
	No justifications are required	Reasoning	TC	
	Work is done only at the formal abstract level	Routine/Drill	TC	
Betty	**Decision:** Task 0 is "rather unsuitable"			
Reasoning	Students first solve very simple multiplication problems	Routine/Drill	TC	**Cluster 3**
		Accessibility	AS	
		Difficulty	AS	
	Students are then led to "backward calculation" (division)	Difficulty	AS	
	Up to problem solving at number 4 (try & combine)	Problem Solving	TC	
	There are different solutions for number 3 on the right side	Openness	AS	
	Rather suitable for low-achieving students	Difficulty	AS	
	No challenge for high-achieving students	Difficulty	AS	

Table 4 The overall mean deviation from expert judgment is aggregated across all eight tasks; a value of 0 represents exact agreement with the experts

	M	*SD*	*t*	*df*	*p*	*d*
Cluster 1	−0.227	0.436	−2.941	31	0.006	0.520
Cluster 2	−0.129	0.392	−1.833	30	0.077	0.329
Cluster 3	−0.117	0.355	−1.273	14	0.224	–

Table 5 Average assessment per cluster in compared to experts

	Task 0	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
Mean expert judgment	4.00	1.67	3.33	4.00	3.00	1.00	1.33	3.67
SD	1.20	0.56	0.61	0.77	0.28	1.00	1.00	0.44
Average assessment of teachers from **Cluster 3**	3.00	1.86	2.80	3.60	3.00	1.79	2.00	3.43
Average assessment of teachers from **Cluster 2**	2.84	2.60	2.73	2.97	2.77	2.10	2.50	3.06
Average assessment of teachers from **Cluster 1**	2.59	1.88	2.65	3.25	2.58	2.07	2.41	3.27

Table 6 Distribution of school type, teaching experience, level of employment, and sex among the three clusters

	Cluster 1	Cluster 2	Cluster 3	X^2	*df*	*p*
School type						
Lower secondary school	23	14	8	4.75	2	0.093
High school	9	17	7			
Teaching experience						
Less or equal to 3 years	3	0	0	8.21	4	0.084
3–9 years	2	7	3			
More than 9 years	12	7	5			
Employment						
Less than 50%	1	2	0	6.74	6	0.346
Exactly 50%	3	0	0			
50–100%	4	2	3			
Exactly 100%	10	10	6			
Sex						
Female	18	16	12	3.82	2	0.148
Male	13	16	3			

shows that some tasks are more difficult for teachers to judge.

3.4 General characteristics of teachers in relation to cluster membership

The descriptive results for RQ4 regarding the relationships among the general characteristics of the teachers and their focus on AS, TC, and TS are presented in Table 6. These results indicate potential differences, which should be interpreted cautiously. A proportion of 51% of lower secondary school teachers belong to Cluster 1—they have a broad view of tasks related to the evaluation of differentiation potential—which differs (marginally) significantly from that of high school teachers, $p = 0.093$; most of them (52%) belong to Cluster 2. One plausible reason for this may be that high school teachers analyze tasks primarily in terms of their mathematical content and not explicitly in terms of features related to their differentiation potential—even if they are asked so. This may also be due to the different teacher education levels in Germany, particularly the stronger subject-specific orientation of teacher education for high school teachers. Teaching experience also showed (marginally) significant differences ($p = 0.064$), with less experienced teachers predominantly found in Cluster 1. However, greater teaching experience does not necessarily guarantee that teachers are found in Clusters 3. Cluster 3 is important for the present study because it focuses broadly on task characteristics that are directly relevant to differentiation (AS); however, the data do not indicate that any one covariate studied may be related to a focus on AS.

Regarding teaching experience (note that only data from 39 teachers are available here; Table 5), the data indicate that teaching experience is (marginal) significantly correlated with belonging to Cluster 3—the strongest cluster, $p = 0.084$. Five teachers with more than nine years of professional experience are found in Cluster 3; there are no newcomers to the profession in this cluster. However, no significant differences regarding either teachers' sex or teachers' level of employment at schools or the composition of the clusters were found (Table 5).

4 Discussion

In this article, we argue that the evaluation and selection of tasks (as a core practice of teaching; Forzani, 2014; Grossman et al., 2009) should be considered relevant components of teaching quality. We assessed how 78 teachers evaluate exercise tasks in terms of their differentiation potential and which task features (and categories) they focus on in these assessments. With the help of a rating scheme and three developed categories, we clustered teachers into three groups according to their focus when evaluating tasks during lesson planning—which can be of general interest for mathematics teacher education.

4.1 Teachers' selection of tasks as a quality criterion for mathematics instruction

One aim of our study is to distinguish between task selection criteria that are *domain-specific* and those that are *domain-general*. We argue that (from the perspective of mathematics education research) task features of the deep structure that are domain-specific should be considered by mathematics teachers to ensure high-quality learning opportunities in their classrooms.

The teachers in our sample are of different sexes, work at different types of schools, have different levels of employment, and have different degrees of teaching experience. However, the data do not indicate an influence on the identification of task features that are relevant for a high differentiation potential (category AS). We thus assume that most of the teachers studied have no experience with adaptive tasks and have not thought systematically about domain-specific task features regarding the differentiating potential of tasks. We were able to identify these teachers by their different focus on task features. Through this differentiated focus, it was possible to discern teachers who likely possessed less experience with adaptive tasks. For example, teachers with limited exposure to teaching mathematics adaptively should not be expected to reference many task features categorized as AS. An illustrative case is Noah, from whose mentioned task features we infer a lack of substantial experience with adaptive tasks.

Specifically, we identified three clusters of teachers. Teachers have different views regarding the differentiation potential of tasks in exercise phases. The category AS comprises *domain-specific* and *directly relevant* task features that are significant regarding the differentiation potential of a task. In *Clusters 1* and *3,* the values for the category AS are increased. Overall, it would be desirable if teachers focused more on features from the category AS. Only 19% of the teachers are in *Cluster 3*—which is closest to an expert-like perspective on task differentiation. From our perspective, we recommend that teachers focus less on task structure and more on what we consider to be a successful combination of task features in Cluster 3. 81% of the teachers do not yet have this specific focus on tasks or still strongly incorporate features of the category TS into their justifications. This finding indicates that most teachers may benefit from further training in adaptive tasks and suggests a demand for a professional development program, affirming its relevance and necessity to enhance teachers' ability to implement differentiated learning effectively.

One reason for the 51% *lower secondary school* teachers in *Cluster 1* (i.e., all three categories are considered in the analysis) could be their experience with different types of tasks, as adaptive tasks can also be found in textbooks for this type of school in Germany (in contrast to high schools). High school teachers account for *domain-specific* and *indirectly relevant* task features, such as language, creativity, and problem solving, and do not consider *domain-specific* and *directly relevant* task features in terms of adaptivity. Instead, they focus on task features from the category TC and the suitability of a task for their own classes as a whole.

The significance of *Cluster 3* in terms of the accuracy of the assessment of the differentiation potential of a task in relation to the expert assessment is high. Thus, teachers from *Cluster 3* are always closer to the experts' value than the average value from other clusters.

We found that teachers focus on drastically different task features when they evaluate mathematical tasks in terms of differentiation potential. We have thus closed a previous research gap: Prior to this study, it was unclear whether teachers could be classified according to the task features they prioritize when selecting a task. Previous studies, when extending beyond case analyses, have shown that teachers generally vary in their ability to identify and use adequate tasks (Hammer & Ufer, 2023; Leuders & Föckler, 2016) and that task features vary in difficulty when being recognized by teachers (Rieu et al., 2022). We can now—by extending such variable-centered approaches to a person-centered approach—posit that teachers have different profiles with respect to utilizing different task features in their evaluations.

Our analysis of the data revealed the existence of three distinct clusters of teachers. Among these, *Cluster 3* stood out as a particularly noteworthy group, as its members consistently identified the specific adaptivity category as a key task feature when making task selection decisions. However, more teacher professional training programs are needed to make the results of our research effective for actual lessons and teaching quality in schools. In German schools, heterogeneous learning groups in mathematics are a common phenomenon. Consequently, mathematics teachers must be prepared to address heterogeneity with tasks. For this to succeed, teachers must be able to independently recognize the potential for differentiation in tasks. It is evident that this knowledge is not yet widespread among teachers. This justifies our call for professional teacher programs.

4.2 Limitations and future research

Overall, the dedicated teachers involved participated in the test voluntarily. It may be that teachers who show less interest in differentiation in the classroom also have a different view of the differentiation potential of tasks. It is possible that teachers with a low interest in differentiation may belong to a cluster that has not yet been identified. This cluster would have a low value in AS and a high value in the category TS. Knowledge about the existence of such a cluster would be relevant, as these teachers require development in terms of switching from their focus on TS to AS.

It is also possible that these teachers focus on other task features. Therefore, theoretically, a new cluster may emerge if these teachers participate in the test—which should be clarified in future studies. In accordance with this, the subsequent correlation of the resulting clusters in future studies with more general aspects of teachers' professional knowledge may yield further insights.

We selected the *differentiation potential* of tasks as one central aspect that should be the focus of mathematics teachers when analyzing tasks during the preparation phase of mathematics instruction. Therefore, we advanced research on task construction (Sullivan et al., 2013) and task implementation (Hiebert et al., 2003b; Stein et al., 1996), and also on task selection (Adleff et al., 2023; Jordan et al., 2008)—but our empirical results are bound to the topic and content investigated and can be applied to different task potentials, e.g., the diagnostic potential of tasks to detect specific student errors.

Furthermore, it appears relevant to ask whether the clusters should be considered rather stable traits (that may be consistent within subjects between different domains of mathematics and over time) or rather likely to be altered states (that may differ within subjects between different domains of mathematics—and may be influenced via short-term interventions). To clarify this, future studies could investigate the stability of those clusters during experimental interventions (e.g., during teacher professional development programs) by comparing different mathematical domains and/or task potentials and by considering teaching experience as a within-subject variable. Therefore, a relevant question not addressed in the present study is the role of teachers' *beliefs* regarding the need to differentiate at all, their teaching goals, and their anticipated learning goals for mathematics classrooms (Anderson, 2003), which could correlate with the clusters found.

One aim of a teacher professional development program should be to train the participants regarding *Cluster 3* (and the task features from the category AS) when focusing on the differentiation potential of tasks—as this may be considered directly related to whether and how teachers react to heterogeneity in their mathematics classrooms. For this purpose, it is highly important to address task features from the deep structure and to sharpen this specific view of tasks by means of specific task examples.

The development of adaptive tasks can also be an important point, as can the use of concrete textbook tasks and the practice of how these tasks can be changed and adapted to achieve high differentiation potential. Moreover, we have not considered implementation in the classroom. This implementation also plays a crucial role in terms of the quality of teaching in terms of differentiation.

The uncertainty of teachers studied regarding adaptive tasks in several empirical investigations—among them the present study—should be a central starting point for developing teacher professional development programs to achieve higher-quality mathematics instruction.

Appendix

Task features related to differentiation potential

Feature	Description
Accessibility	The teacher states that the task contains introductory parts that enable successful entry, in particular to those with lower abilities (or learners in general) (for example, understanding the task, simple cases, concrete examples)
Change of mathematical representation	The teacher states that the task requires a change of presentation, a transfer of fractional representation into "normal" language use or into a different notation
Context/application	The teacher states that the task contains an application reference, illustrative content, (everyday) context, or is abstract
Creativity	The teacher states that the task stimulates learners' creativity (own ideas, concepts, objects, …)
Difficulty	The teacher describes the level of difficulty/level of the task or whether the learners are overwhelmed
Generalizing	The teacher states that the task requires learners to recognize patterns and to generate or generalize conjectures or questions
Goal differentiation	The teacher states that the task pursues different content goals for high-achieving and low-achieving students
Guidance	The teacher states that the task provides guidelines (e.g., possible outcomes) or a pattern for approaching the work

Feature	Description
Language	The teacher states that the task is simple in its language (that is, it does not contain any educational or technical language requirements or formal spellings) and is understandable even for low-achieving students
Layout	The teacher describes the (small step) presentation/arrangement/editing or the format of the illustrations, parts of text, and so on of the task
Miscellaneous	For example, the task is well suited for partner work; independence is needed
Motivation	The teacher states that the learners are motivated by the task, that the task is varied, or that frustration tolerance is high
Openness	The teacher describes the openness of the (sub)task, which allows for different/individual solutions/approaches or results (problem solving required)
Prerequisites	The teacher states that the task requires learners to have prior knowledge/prerequisites for processing—for example, computer security, background knowledge, presentation of the size of fractures
Presentation	The teacher states that the task contains representations or content that is unusual and potentially confusing to the learner
Problem solving	The teacher states that the task requires learners to try something in depth, to do many things at once, to solve problems, to puzzle over, or to make more open/complex decisions
Reasoning	The teacher states that the task requires an explanation of/reflection on a procedure or an argumentation from the learner
Routine/drill	The teacher states that the task requires the application/reproduction of known (computational) procedures; it is all about automation, not understanding ("Learning task," development would be)
Self-checking	The teacher states that the task is designed to encourage learners' independence
Size	The teacher describes the scope/processing time of the task, the number of tasks, or that many types of content are present in the task
Support	The teacher states that the task provides an additional (substantial) pedagogic aid, such as an explanation, a concrete example, or a visualization (picture/drawing)
Transparent structure	The teacher states that the approach to working on the (sub) task(s) is (completely) clear, confusing, or complex to the learner

Acknowledgements We would like to thank the German Centre for Mathematics Teacher Education (DZLM), funded by the Deutsche Telekom Foundation, for supporting this study. In addition, we would like to thank the Ministry of Education in Saarland and the State Education Offices in Freiburg, Offenburg, and Böblingen for giving us opportunities to conduct our professional development course.

Funding Open access funding provided by FHNW University of Applied Sciences and Arts Northwestern Switzerland.

Declarations

Conflict of interest All the authors declare that they have no conflicts of interest.

References

Adleff, A. K., Ross, N., König, J., & Kaiser, G. (2023). Types of mathematical tasks in lower secondary classrooms in Germany: Statistical findings from a latent class analysis based on general mathematical competencies. *Educational Studies in Mathematics, 114*(3), 371–392. https://doi.org/10.1007/s10649-023-10254-9

Anderson, J. (1996). Some teachers' beliefs and perceptions of problem solving. In P. C. Clarkson (Ed.), *Technology in mathematics*

education (Proceedings of the 19th Conference of the Mathematics Education Research Group of Australasia) (pp. 30–37). MERGA.
Anderson, J. (2003). Teachers' choice of tasks: A window into beliefs about the role of problem solving in learning mathematics. In L. Bragg, C. Campbell, G. Herbert, & J. Mousley (Eds.), *Mathematics educations research: Innovation, networking, opportunity (Proceedings of the 26th Conference of the Mathematics Education Research Group of Australasia)* (pp. 72–79). MERGA.
Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433–456). Macmillan.
Bardy, T., Holzäpfel, L., & Leuders, T. (2021). Adaptive tasks as a differentiation strategy in the mathematics classroom—Features from research and teachers' views. *Mathematics Teacher Education and Development, 23*(3), 26–53.
Blömeke, S., Risse, J., Müller, C., Eichler, D., & Schulz, W. (2006). Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht: Ein allgemeines Modell und seine exemplarische Umsetzung im Unterrichtsfach Mathematik. *Unterrichtswissenschaft, 34*(4), 330–357.
Bölsterli Bardy, K., & Wilhelm, M. (2018). Von kompetenzorientierten zu kompetenzfördernden Aufgaben im Schulbuch. *Erziehung und Unterricht, 168*(1/2), 121–129.
Brod, G. (2024). There are multiple paths to personalized education, and they should be combined. *Current Directions in Psychological Science, 33*(3), 153–158. https://doi.org/10.1177/09637214241242459
Bromme, R. (1981). *Das Denken von Lehrern bei der Unterrichtsvorbereitung: Eine empirische Untersuchung zu kognitiven Prozessen von Mathematiklehrern*. Beltz.
Chapman, O. (2013). Mathematical-task knowledge for teaching. *Journal of Mathematics Teacher Education, 16*(1), 1–6. https://doi.org/10.1007/s10857-013-9234-7
Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software, 61*, 1–36. https://doi.org/10.18637/jss.v061.i06
Dreher, A. (2015). *Dealing with multiple representations in the mathematics classroom—Teachers' knowledge, views, and their noticing*. Doctoral dissertation, Pädagogische Hochschule Ludwigsburg. https://phbl-opus.phlb.de/frontdoor/deliver/index/docId/61/file/Thesis+Anika+Dreher.pdf
Dreher, A., & Leuders, T. (2021). Fachspezifität von Unterrichtsqualität—aus der Perspektive der Mathematikdidaktik. *Unterrichtswissenschaft, 49*(2), 285–292. https://doi.org/10.1007/s42010-021-00116-9
Forzani, F. M. (2014). Understanding "Core Practices" and "Practice-Based" teacher education: Learning from the past. *Journal of Teacher Education, 65*(4), 357–368. https://doi.org/10.1177/0022487114533800
Gallagher, M. A., Parsons, S. A., & Vaughn, M. (2022). Adaptive teaching in mathematics: A review of the literature. *Educational Review, 74*(2), 298–320. https://doi.org/10.1080/00131911.2020.1722065
Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching, 15*(2), 273–289. https://doi.org/10.1080/13540600902875340
Hammer, S. (2016). *Professionelle Kompetenz von Mathematiklehrkräften im Umgang mit Aufgaben in der Unterrichtsplanung: Theoretische Grundlegung und empirische Untersuchung*. Doctoral dissertation, LMU Munich. https://edoc.ub.uni-muenchen.de/20439/1/Hammer_Sabine.pdf
Hammer, S., & Ufer, S. (2023). Professional competence of mathematics teachers in dealing with tasks in lesson planning. *Teaching and Teacher Education*. https://doi.org/10.1016/j.tate.2023.104246
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
Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., & Jacobs, J., et al. (2003a). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study (NCES 2003-013)*. National Center for Education Statistics. https://nces.ed.gov/pubs2003/2003013.pdf
Hiebert, J., Morris, A. K., & Glass, B. (2003b). Learning to learn to teach: An "Experiment" model for teaching and teacher preparation in mathematics. *Journal of Mathematics Teacher Education, 6*, 201–222. https://doi.org/10.1023/A:1025162108648
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., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., et al. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht. *Journal Für Mathematik-Didaktik, 29*(2), 83–107. https://doi.org/10.1007/BF03339055
Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., & Löwen, K., et al. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenklassifikation im COACTIV-Projekt*. Max-Planck-Institut für Bildungsforschung. https://pure.mpg.de/rest/items/item_2100753_3/component/file_2197661/content
Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM - the International Journal on Mathematics Education, 40*(5), 873–892. https://doi.org/10.1007/s11858-008-0141-9
Kuntze, S. (2011). In-service and prospective teachers' views about modelling tasks in the mathematics classroom—Results of a quantitative empirical study. In G. Kaiser, W. Blum, R. Borromeo Ferri, & G. Stillman (Eds.), *Trends in teaching and learning of mathematical modelling. International perspectives on the teaching and learning of mathematical modelling* (Vol. 1, pp. 279–288). Springer. https://doi.org/10.1007/978-94-007-0910-2_28
Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718. https://doi.org/10.3102/0034654315627366
Leuders, T., & Föckler, F. (2016). Aufgabenqualität im Fach Mathematik. Differenzierungsvermögen und kognitive Aktivierung. In T. Bohl & A. Wacker (Eds.), *Die Einführung der Gemeinschaftsschule in Baden-Württemberg. Abschlussbericht der Wissenschaftlichen Begleitforschung (WissGem)* (pp. 213–226). Waxmann.
Leuders, T., & Prediger, S. (2016). *Flexibel differenzieren und fokussiert fördern im Mathematikunterricht*. Cornelsen.
Maier, U., Kleinknecht, M., Metz, K., & Bohl, T. (2010). Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben. *Beiträge Zur Lehrerinnen- und Lehrerbildung, 28*(1), 84–96. https://doi.org/10.36950/bzl.28.1.2010.9798
Morris, A. K., Hiebert, J., & Spitzer, S. M. (2009). Mathematical knowledge for teaching in planning and evaluating instruction: What can preservice teachers learn? *Journal for Research in Mathematics Education, 40*(5), 491–529. http://www.jstor.org/stable/40539354
Neubrand, J. (2006). The TIMSS 1995 and 1999 video studies. In F. K. S. Leung, K. D. Graf, & F. J. Lopez-Real (Eds.), *Mathematics*

education in different cultural traditions—A comparative study of East Asia and the West: The 13th ICMI study (pp. 291–318). Springer. https://doi.org/10.1007/0-387-29723-5_18

Ostermann, A., Leuders, T., & Nückles, M. (2018). Improving the judgment of task difficulties: Prospective teachers' diagnostic competence in the area of functions and graphs. *Journal of Mathematics Teacher Education, 21*(6), 579–605. https://doi.org/10.1007/s10857-017-9369-z

Philipp, K. (2018). Diagnostic competences of mathematics teachers with a view to processes and knowledge resources. In T. Leuders, K. Philipp, & J. Leuders (Eds.), *Mathematics teacher education: Vol. 11. Diagnostic competence of mathematics teachers: Unpacking a complex construct in teacher education and teacher practice* (pp. 109–127). Springer. https://doi.org/10.1007/978-3-319-66327-2_6

Pilous, R., Leuders, T., & Rüede, C. (2023). Novice and expert teachers' use of content-related knowledge during pedagogical reasoning. *Teaching and Teacher Education*. https://doi.org/10.1016/j.tate.2023.104149

Praetorius, A.-K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM Mathematics Education, 50*(3), 535–553. https://doi.org/10.1007/s11858-018-0946-0

Rieu, A., Leuders, T., & Loibl, K. (2022). Teachers' diagnostic judgments on tasks as information processing—The role of pedagogical content knowledge for task diagnosis. *Teaching and Teacher Education*. https://doi.org/10.1016/j.tate.2021.103621

Seidel, T., Blomberg, G., & Stürmer, K. (2010). "Observer"–Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. *Zeitschrift Für Pädagogik, 56*(Beiheft), 296–306.

Smith, M. S., & Stein, M. K. (1998). Reflections on practice: Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School, 3*(5), 344–350. https://doi.org/10.5951/MTMS.3.5.0344

Stein, M. K., Grover, B., & 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/10.2307/1163292

Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection: From research to practice. *Mathematics Teaching in the Middle School, 3*(4), 268–275. https://doi.org/10.5951/MTMS.3.4.0268

Stillman, G. (2004). Strategies employed by upper secondary students for overcoming or exploiting conditions affecting accessibility of applications tasks. *Mathematics Education Research Journal, 16*(1), 41–71. https://doi.org/10.1007/BF03217390

Sullivan, P. (1999). Seeking a rationale for particular classroom tasks and activities. In J. M. Truran, & K. N. Truran (Eds.), *Making the difference (Proceedings of the 21st Annual Conference of the Mathematics Educational Research Group of Australasia)* (pp. 15–29). MERGA.

Sullivan, P., Clarke, D., & Clarke, B. (2013). *Teaching with tasks for effective mathematics learning*. Springer. https://doi.org/10.1007/978-1-4614-4681-1

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

van Es, E., & Sherin, M. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education, 10*(4), 571–596.

Wellenreuther, M. (2004). *Lehren und Lernen–aber wie? Empirisch-experimentelle Forschung zum Lehren und Lernen im Unterricht*. Schneider Hohengehren.

Wessel, L., & Erath, K. (2018). Theoretical frameworks for designing and analyzing language-responsive mathematics teaching–learning arrangements. *ZDM Mathematics Education, 50*(6), 1053–1064. https://doi.org/10.1007/s11858-018-0980-y. Special Issue: Making frameworks for the study of language in mathematics education research visible.

Yeo, J. B. W. (2017). Development of a framework to characterise the openness of mathematical tasks. *International Journal of Science and Mathematics Education, 15*(1), 175–191. https://doi.org/10.1007/s10763-015-9675-9

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