Adaptive Tasks as a Differentiation Strategy in the Mathematics Classroom: Features from Research and Teachers' Views

Thomas Bardy
University of Education Freiburg

Lars Holzäpfel
University of Education Freiburg

Timo Leuders
University of Education Freiburg

Received: 5th February, 2021/ Accepted: 26th April, 2021

Tasks play a central role in the mathematics classroom. Especially when teaching in a heterogeneous mathematics class, teachers should be able to find, select, modify, and assign tasks adequately. The focus in this paper is on adaptivity to cognitively activate all learners at the individual level, and on teachers' abilities to allow for such adaptivity by means of selecting appropriate tasks. More specifically, when planning lesson phases for practice and consolidation, teachers may consider which tasks have differentiation potential and can thus be completed by all students at the same time but at different levels. In order to analyse teachers' strategies in this regard, we investigated the task features that teachers may consider when assessing the differentiation potential of exercise tasks. We first deductively constructed rating categories based on the literature on instructionally relevant features of adaptive tasks. Then, we inductively extended and refined the constructed categories by analysing teachers' reasoning based on a sample of 78 in-service teachers at secondary schools. We validated the resulting 22 categories by determining interrater reliability. Our findings indicate that teachers consider a broad spectrum of task features when analysing the differentiation potential of tasks. However, only some of these features are directly relevant with regard to using adaptive tasks as a differentiation strategy. Our results also show that many teachers arrived at conclusions about the differentiation potential of tasks that were different from task-design experts. Based on our findings regarding teachers' perspectives on the differentiation potential of tasks and on certain task features, we discuss how these findings may have arisen and how important the knowledge about the deep structure of adaptive tasks is for teachers' professional development.

Keywords · differentiation · adaptivity · adaptive tasks · task features · teacher professional development

Introduction

Strategies for teaching in a heterogeneous classroom have been an important topic in education research and practice for a long time (e.g., Pozas et al., 2020). The umbrella terms "differentiated instruction," or "adaptive instruction" cover a multiplicity of such strategies, including organization level strategies (e.g., ability grouping; Lou et al., 1996), general instructional formats (e.g., diagnostic teaching; Brown, 1984; adaptive teaching; Beck et al., 2008), and domain- or topic-specific strategies (e.g., task design; Barzel et al., 2013; Clarke et al., 2014; Leuders & Prediger, 2016; using prompts; Russo et al., 2020), which teachers often combine in practice. As teacher professional development is more effective when it is focused on student thinking and on specific topics and situations in practice (e.g., Lipowsky & Rzejak, 2015), there is a need for research that generates knowledge on specific teaching practices and specific teacher knowledge and learning (e.g., Carpenter et al., 1999).

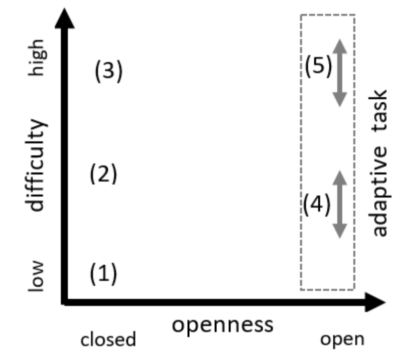
In our study, we investigate teachers' perspectives when assessing tasks for differentiating instruction. We focus on the suitability of tasks for practice phases in heterogeneous learning groups. We believe the focus on tasks to support differentiated instruction is justifiable as there is substantial evidence that tasks have a crucial impact on students' thinking and understanding when learning mathematics (e.g., Doyle, 1983, 1988; Bromme et al., 1990; Hiebert & Wearne, 1997; Watson & Mason, 2006; Zaslavsky & Sullivan, 2011; Chapman, 2013; Leuders, 2015; Hammer, 2016; Sullivan et al., 2016; Russo et al., 2020).

There are numerous ways to design differentiated instruction. One approach is to assign different tasks to different students and/or provide different types of support to different student groups (here, we are referring to "closed differentiation," whereby the teacher facilitates adaptivity by providing appropriate tasks; Snow, 1989, pp. 605–606; Helmke, 2010, p. 247). In a heterogeneous classroom, a teacher's discursive skills in orchestrating class discussion on the solution of tasks is also considered highly important (e.g., Doyle, 1988; Beck et al., 2008). Beyond these "closed" and "discursive" approaches, there is also an "open" approach to differentiation, which is at the heart of our theoretical and empirical analysis: we specifically focus on a differentiation strategy that uses identical tasks for all students in a class. We call these tasks "adaptive tasks" because adaptivity unfolds as students work on a task in practice and consolidation phases.

In the literature, the type of task required for this purpose is referred to as "adaptive task," "open adaptive task," "open-ended task," "self-differentiating task," or "open differentiating task" (e.g., Müller & Wittmann, 1998; Sullivan, 1999). Scholars have researched the features and instructional uses of this task in different contexts (e.g., Boston & Smith, 2009; Prediger & Scherres, 2012; Leuders & Prediger, 2016). An "adaptive task" is a question or an activity that contains multiple starting points and solutions for students of different levels so that every student can participate. Based on Vygotsky's theory of the zone of proximal development (Vygotsky, 1978), the idea is to cognitively activate all students at their respective ability levels using the same task. Each student is then challenged to work on the adaptive task at his or her own level of performance. These task types are designed to be used in practice phases (Wittmann & Müller, 1990; Sullivan et al., 2009; Prediger et al., 2021), and students can decide for themselves at which level they want to work.

When it comes to evaluating tasks (generally, and with respect to adaptivity), it is important to systematically consider task features (see Doyle, 1988; Hammer, 2016; Leuders & Prediger, 2016). In this study, we focus on the features of "openness," which is central when dealing with heterogeneity, and "difficulty," which we examine in relation to openness (see Figure 1). The goal of the tasks we consider is to cognitively activate students with different abilities at their respective levels in their proximal development. Figure 1 illustrates how this goal can be achieved within the structure of a task.

In Figure 1, Task (1) is especially suitable for low-achieving students, while Task (2) is more difficult because the denominators are different. Even though this task is more difficult than the previous one, low-achieving students may still be able to solve it. Task (3) is even more complex because finding the common denominator is harder than in the previous two tasks. The "difficulty" feature of a task refers to the number of steps and links that students are expected to make in order to identify a solution. However, none of the Tasks (1) to (3) are adaptive in terms of difficulty. Tasks (4) and (5) are different in this regard. The two tasks demonstrate the feature of "openness," which allows students to work at different levels and enables flexible solution finding. A task can be said to possess the feature of "openness" if its solution path or goal is not predetermined—that is, if students can make their own decisions when working on the task.



Example-Task	Description
(1) $\frac{1}{5}+\frac{2}{5}=$	same denominator
(2) $\frac{1}{3}+\frac{1}{6}=$	easy to find common denominator
(3) $\frac{3}{17}+\frac{5}{13}=$	complicated to find common denominator
(4) Find different solutions: $\frac{\square}{\square}+\frac{\square}{\square}=\frac{\square}{\square}$	free choice, it does not have to be difficult
(5) Find different solutions: $\frac{\square}{\square}+\frac{\square}{\square}=\frac{2}{7}$	free choice, but high difficulty after finding $\frac{1}{7}+\frac{1}{7}$

Figure 1. Tasks in practice phases can be categorized with respect to their varying level of "openness" and "difficulty".

Tasks (4) and (5) are "adaptive tasks." Task (4) encourages students to find many different solutions. Here, not only the scope but also the "difficulty" level is open. Thus, accessibility to the task is ensured even for the low-performing students (Sullivan et al., 2009). Task (5) is a rather complex assignment because only the result is provided. One reason for using adaptive tasks in practice phases is to meet the needs of learners with different skill levels in a single mathematics class.

The task in Figure 2, which was part of our questionnaire (see Method), illustrates a different way of achieving adaptivity in tasks. In this task, two different levels, one for low-achieving and one for high-achieving students, are separated by a vertical dashed line. This is a special task type, in which differentiation is guided by the "parallel" surface structure. The deep structure in this parallel task is characterized by the features "difficulty" and "openness." We call this a "parallel adaptive task" because the left-hand side with the Subtasks (1) to (4) is designed for low-achieving students and the right-hand side with the Subtasks (1) to (4) is designed for high-achieving students. Moderately-achieving students can choose either side. In principle, it is also possible for students to change sides while working on the task. Overall, a teacher should take care that, for example, the high-achieving students are not under challenged when voluntarily choosing the left-hand side of the parallel adaptive task. If students choose this side only because of convenience or effort avoidance, the teacher must intervene. Choosing the right "task side" is also a longer-term learning process for students if they choose a side that is appropriate for them.

Furthermore, on the left-hand side, the Subtasks (1) to (4) exhibit increasing difficulty, which is comparable to the Tasks (1) to (3) in Figure 1. On the right-hand side, the Subtasks (1) to (3) are more difficult than the subtasks on the left-hand side, and the Subtask (4) on the right-hand side is open because it has more than one solution (comparable to Task (4) in Figure 1). Subtask (4) on the left-hand side is not open. High-achieving students know from their teacher's introduction to such task types that they should also find multiple solutions in an open task such as Subtask (4). The teacher expects high-achieving students to recognize that this subtask allows for multiple solutions. By working on the left- or the right-hand side, students practice the same mathematical content but at different levels. At this point, it must be mentioned that, in this case, we understand a task as a set of subtasks that are processed as a whole (in this case, e.g., the four subtasks on the left- and right-hand sides).

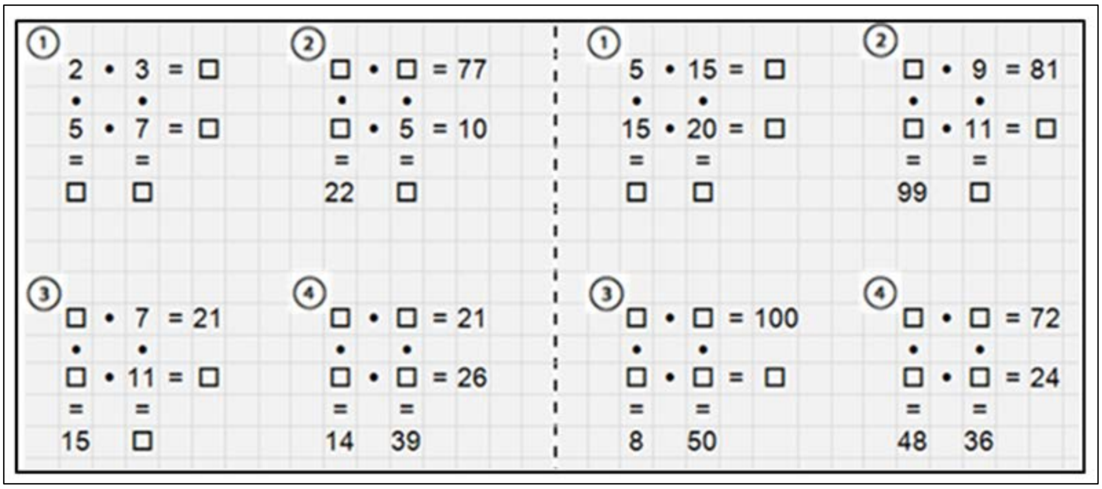


Figure 2. An adaptive task structured as a "parallel adaptive task" for low-achieving (left-hand side) and high-achieving (right-hand side) students (Leuders & Prediger, 2016, p. 121; Task 0 from our questionnaire).

These considerations regarding adaptive task design are useful to teachers who plan to address heterogeneity in mathematics classes by using adaptive tasks to cognitively activate all learners at their individual levels. However, little systematic research has been conducted on teachers' knowledge of tasks in general (e.g., Stein et al., 1996), and specifically on task features as discussed here.

Our Research Interest

Having considered the importance of tasks in mathematics education (e.g., Doyle, 1988; Schoenfeld, 1988), it is now important to analyse, especially with regard to heterogeneous classes, what competencies teachers already have for recognizing the tasks that are potentially suitable for supporting differentiation. Hammer (2016) has already carried out valuable preliminary work on this topic, by analysing teachers' strategies in identifying task features and selecting tasks, although she does not specifically address the issue of dealing with heterogeneity. Therefore, we build on Hammer's preliminary work and focus on the differentiation potential of tasks, with an emphasis on the special features of "openness" and "difficulty." By the "differentiation potential" of a task, we are referring to the implementation of concrete adaptive task features in the deep structure of a task (e.g., the specific adaptive features that can be found in a task). Our goal is to investigate teachers' competencies in this matter and to contribute to the development of teacher training programs (Prediger et al., 2019) with a special focus on the topic of adaptive tasks. Our aim is to examine teachers' reasoning when considering the differentiation potential of tasks.

To this end, we developed a questionnaire to evaluate tasks in terms of their differentiation potential in practice phases. This evaluation addresses a teacher's central activity, namely the selection of suitable tasks for lesson planning (see Hammer, 2016). More specifically, the reasons that teachers give for choosing or not choosing a certain adaptive task indicate the teachers' competencies in recognizing the differentiation potential of tasks. In the next section of this paper, we examine teachers' perspectives on dealing with adaptive tasks and discuss previous research on this matter. However, we would like to acknowledge that while tasks alone are not enough to implement successful differentiated instruction in practice phases, they are a fundamental prerequisite. The actual implementation of adaptive tasks in the classroom is also crucial, but we do not have the space to discuss this topic here.

Reviewing the Literature on Adaptive Tasks

As stated in the Introduction, this paper focuses on the following questions: Which task features are important for adaptive tasks, and to what extent do teachers recognize these features when considering and reasoning the suitability of tasks for students in heterogeneous classes? In other words, we are concerned with the teachers' view on the adaptivity of tasks, and therefore in this section, we will systematically describe in more detail the task features that enable adaptivity.

In the education theory literature, there is a large body of research on task characteristics, although these do not specifically focus on the aspect of differentiation (e.g., Anderson, 1996; Stein et al., 1996; Stein & Smith, 1998; Wellenreuther, 2004; Blömeke et al., 2006; Doyle, 1988; an exception is Leuders & Prediger, 2016). Figure 3 shows an overview of selected literature, which illustrates task features in the literature. The literature references mentioned in Figure 3 each have a different focus on tasks. For example, Anderson (1996) focuses on the "openness" of tasks, Stein et al. (1996) and Maier et al. (2010) on the "cognitive requirements" of tasks and Bölsterli Bardy and Wilhelm (2018) analyse tasks with regard to students' competencies that can be promoted by processing tasks. Depending on the focus a researcher has on tasks, he or she looks at different task features.

	Anderson (1996)	Stein et al. (1996)	Stein & Smith (1998)	Sullivan (1999)	Neubrand (2002)	McDougall (2004)	Wellenreuther (2004)	Jordan et al. (2006)	Blömeke et al. (2006)	Maier et al. (2010)	Leuders & Prediger (2016)	Hammer (2016)	Bölsterli Bardy & Wilhelm (2018)
Focus / Features	Com-ments on open-ness	Cognitive require-ments	Cognitive require-ments	Open-ended tasks	Analysis of classroom situations	„Rich mathe-matical task"	Differentiating task characteristics or contexts	Classifi-cation scheme at COACTIV	Task quality for learning process-stimulating tasks	Analysis of the cognitive potential of tasks	Deep struc-ture	Cognitive demand level	Com-petence-promoting task sets
Openness	x			x	x	x	x	x	x	x	x		x
Difficulty (Complexity)			x	x			x	x	x		x	x	
Language	x				x		x	x	x	x	x		
Help settings (Guidance/ support)		x	x	x	x	x		x	x			x	x
Accessibility				x							x		

Figure 3. Task features in (selected) literature (translated into English).

In Figure 3 we list five features from the set of task features we found in the literature. These features are important for us with respect to an analysis of the differentiation potential of a task and a starting point of our analyses as a whole. The features "openness" and "difficulty" have been analysed in detail in the Introduction (see Figure 1) with respect to the differentiation potential of a task; whereby we classify the task feature "complexity" from the literature under the feature "difficulty".

"Language" is also an important feature with regard to the differentiation potential of a task. A differentiating linguistic formulation in a task can support low- and high-achieving students, for example, to understand the task. The feature "help settings" can be realized in a task to help especially low-achieving students to complete the task. Other features mentioned in the literature such as "guidance" and "support" have the same goal.

In the following section, we examine the task features that focus on the deep structure of tasks in relation to different student achievement levels. The surface structure of tasks, which, for example, allows selecting simpler or more complex subtasks, is also considered in this review.

First, we consider the deep structure of tasks. More specifically, we focus on the dimensions of "openness" and "difficulty" discussed earlier. Afterwards, we take up the surface structure of tasks, which, by means of task presentation, allows learners to choose from different options. Later we focus on teachers' perspectives on adaptive tasks as described in selected literature.

The Deep Structure of Tasks

Central to this section is the issue of how to stimulate cognitive activation in all learners. Our starting point is the consideration that all students should be cognitively activated in their zones of proximal development (see Vygotsky, 1978). In this respect, a certain "openness" in the tasks is essential to encourage the involvement of individual students. Not infrequently, the problem is that only a few students are offered the optimal adaptation of course material (e.g., tasks) to their level. Rather, what happens matches the following description by Doyle (1988): "Although students often accomplish a large amount of work, they seldom appear to be faced with tasks in which they are required to struggle with meaning" (Doyle, 1988, p. 177).

So, what can be done to ensure that tasks offer more flexibility in terms of what is required of students? What features do tasks need to have to challenge students in their individual zones of proximal development? In a study from Baumert, Blum and Neubrand on cognitive activation (COACTIV) in the classroom (e.g., Kunter et al., 2013), tasks were classified by the research group in terms of cognitive activities, taking into account different facets of knowledge, such as procedural and conceptual knowledge, and different levels of difficulty (Jordan et al., 2008, p. 88). Doyle (1983) also approached cognitive processes as relevant for the activation of students, and emphasized the difference between higher and lower cognitive processes.

We also start our considerations at the level of cognitive activities and ask to what extent a task can accommodate learners with different achievement levels. The task feature of "openness" is well suited for encouraging task adaptability because it allows the students to choose the level at which they want to or can deal with a problem. However, as Prediger and Scherres (2012) demonstrate, it is by no means guaranteed that students choose a level of working with tasks that corresponds ideally with their ability. Consequently, sub-optimal outcomes have to be envisaged when applying adaptive tasks with respect to the match between the level of challenge experienced and the ability level of the learner.

Difficulty of Tasks

Blömeke et al. (2006, p. 347) noted that differentiation potential means that tasks can be worked on at different depths and at different cognitive levels and that students can decide on how to proceed. In this sense, the cognitive processes activated as students work towards a solution are of central importance to identify the potential for differentiation. In the example that we presented earlier, the difficulty varies according to one's choice of fractions with equal or unequal denominators and simpler or more difficult numbers (see the gradation of Tasks (1), (2), (3) in

Figure 1). A certain level of difficulty in tasks is necessary to trigger processes that are critical to learning.

Schoenfeld (1988) reported on a study in which he found many small-step tasks in several classes over an entire school year, with few tasks being really challenging: "What students worked were exercises: tasks designed to indicate mastery of relatively small chunks of subject matter, and to be completed in a short amount of time" (p. 13). What we learn from these studies is that while it is important that tasks remain accessible to low-achieving students, higher levels of cognitive stimulation are needed to engage high-achieving students. Thus, there is a challenge in using an adaptive task to provide practice opportunities to low-achieving students and high-achieving students. But which task features are important for this purpose? To answer this question, a look at the research work of the KOSIMA-project (Prediger et al., 2021) may be helpful. The project investigates multiple aspects of mathematical learning processes in meaningful contexts. In particular, in this project tasks were developed and tested from the researchers that have precisely the goal of promoting differently performing students. These tasks (see Figure 2 and 4) have features that describe a high level of differentiation potential.

The example tasks in Figure 1 show that the task feature "difficulty" (e.g., Wellenreuther, 2004; Leuders & Prediger, 2016; see Figure 3) gradually increased from Task (1) to Task (3). The task feature of "difficulty" involves and summarizes several characteristics that result in the difficulty of a given task such as complexity or formalization (e.g., Stein & Smith, 1998; Blömeke et al., 2006; Hammer, 2016). Leuders and Prediger (2016) distinguished between the task features "difficulty" and "complexity." According to this distinction, the feature "complexity" describes the technical complexity involved in processing a specific task—for example, due to the choice of large numbers, demanding number ranges, or complicated terms.

Openness of Tasks

The second dimension, which is important in relation to the variation of "difficulty," is the "openness" of tasks, as only by means of openness can the different difficulty levels be variably offered and selected by the students. The "openness" of a task describes the extent of the decisions that students can make regarding the processing method or the goal. For example, in Task (4) (see Figure 1), a student can insert a fraction, then insert the next fraction, and then determine the result of the addition. Alternatively, a student can start from one result on the right-hand side and determine the two missing addends. Overall, we describe this task as "open" with respect to different ways of solving this task and its different solutions. Of course, the various paths toward a solution are not arbitrary, but "openness" leaves a certain degree of freedom to students. How can the openness of a task be designed to include different performance levels?

Before tackling the question of "openness," we first had to deal with the different difficulty levels of tasks, as in the section above, as including varying difficulty levels is fundamental to the idea that students can choose different levels of difficulty themselves. The importance of open-ended tasks for challenging all learners was also mentioned by Anderson (1996; see Figure 3), who stated the following: "Open-ended problems have several solutions with potentially different ways of finding and recording solutions" (p. 73).

The "openness" (e.g., Wellenreuther, 2004) of Tasks (4) and (5) in Figure 1 allowed both high- and low-achieving students to work on them. In Task (4) low-achieving students can find different solutions to the gaps by substituting fractions from the left-hand side to the right-hand side. High-achieving students can also find solutions, but they can find more complicated fractions. In Task (5) in Figure 1, low-achieving students can find at least one solution. High-achieving students will be able to determine several different solutions in this task.

Tasks can be analysed not only from their deep structure but also from their surface structure. The surface structure of a task can also, like the deep structure be an indication of the suitability of the task for differentiated use in the classroom. Therefore, we describe different surface structures of mathematics tasks in the next section.

The Surface Structure of Tasks

In this section, we consider the external features of a task, as elaborated by Leuders and Prediger (2016): the "surface structure" differs from the "deep structure" described in the previous section and deals with the presentation of a task, which enables navigation in the processing of a task. This makes it easier for students to decide what exactly they would like to work on and in what way, depending on their performance level.

The surface structure of a task gives teachers and students a visual indication of the different difficulty levels. Such external indications involve, for example, a parallel representation of two task sets, with one set being easier and the other more difficult to process, as shown in Figure 2. In this figure, similar mathematical content is processed on both sides, but the tasks involve different characteristics (in Figure 2, e.g., tasks on the left-hand side have lower complexity, smaller numbers, the subtasks here have only one solution). Low-achieving students would work on the left-hand side of the task, high-achieving students on the right-hand side, and moderately-achieving students could choose or switch between the two (e.g., in Figure 2, from (1) to (3) on the right-hand side to (4) on the left-hand side). We call this type of task a "parallel adaptive task."

Add the numbers. The result is always in the middle of two numbers. Search for several solutions in (D).

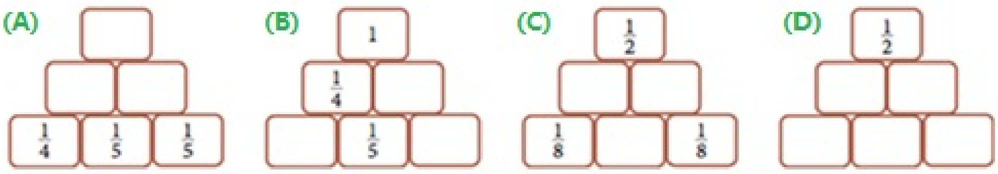


Figure 4. Graded adaptive task (Leuders & Prediger, 2016, p. 121, translated into English; this was Task 3 in our questionnaire).

However, an adaptive task can also be designed as a "graded adaptive task," in which one can (but does not have to) start at a very simple entry point in the first subtask and then move, step by step, to a higher level of difficulty (see Figure 4). In our example, low-achieving students will find a simple starting point with Subtask (A); high-achieving students can, for example, start with Subtask (C). The openness of Subtask (D) and the difficulty of (C) should also be noted.

Here you have 24 wooden cubes.
Which cuboids can you build with them?
Make a note of the ones you have already found.
How many can you find?

Figure 5. A self-adaptive task (Leuders & Prediger, 2016, p. 15, translated into English).

"Self-adaptive tasks" combine several of the task characteristics already discussed. What is special about this task type is its suitability for both low- and high-achieving students. All students work on the same task at different levels and different speeds and using different approaches. Figure 5 provides an example of a self-adaptive task.

This task type has been discussed in the literature—for example, by Heymann (1991), Müller and Wittmann (1998), Hirt and Wälti (2008), and Krauthausen and Scherer (2010). The main features of this task type are openness and different difficulty levels, approaches, and ways of doing things (e.g., lay or draw cubes; see Hengartner et al., 2006; Hußmann & Prediger, 2007; Leuders & Prediger, 2016).

Scherres (2012) showed that in certain self-adaptive tasks with a high degree of openness, optimal learning often requires that the learning process is supported by a teacher intervention that transforms the mathematical learning potential into a fruitful epistemological situation. As we are looking at self-adaptive tasks in exercise phases, it can be assumed that only low-achieving students need teacher support. A teacher expects their high-achieving students to find a solution without help during an exercise phase.

Teachers' Perspectives on Adaptive Tasks

In this paper, we focus on teachers' abilities to recognize the differentiation potential of tasks, which is a prerequisite for adapting course material. Hammer's (2016) work examined quite generally the task features and potentials that teachers consider when planning instruction. By focusing on the issue of adaptivity, we continue this line of inquiry.

We assume that the first and necessary step in the selection of tasks for the differentiating practice phases is for teachers to recognize the task features that potentially support differentiated instruction. Therefore, we focus on the adaptive tasks that can be worked on by students of different levels at the same time—that is, there is no need to group students or assign tasks of different levels. The task features that are relevant to adaptive tasks refer either to the surface structure (e.g., parallel adaptive tasks or graded adaptive tasks) or to the deep structure (e.g., difficulty, openness) of a task.

Some studies provide teachers' comments on the use of, for example, "open tasks" and categorize these comments (Anderson, 2003). However, these comments do not consider the differentiation potential of tasks. This is precisely the focus of our research, namely, how teachers assess the potential of tasks in terms of adaptive use (i.e., the learning levels of different students) in practice phases and thus what knowledge they have about task features in terms of differentiation potential.

A preliminary study by Hammer (2016) examined whether and to what extent professional perception (e.g., van Es & Sherin, 2002; Seidel et al., 2010) is, in principle, suitable for examining task handling in lesson planning. Among other things, Hammer also examined the features that teachers referred to when justifying their task selection.

Hammer's study was the first attempt to understand teachers' reasoning regarding task selection on qualitatively different levels. Teachers justified their task selection using the following concepts: cognitive activation, constructive learning support, networking of mathematical ideas, learning objectives from educational standards, or specific teaching activities, such as "to practice." The "quality" of justification of the selection was determined by the extent to which the teachers referred to the students' learning process. In the "Coding Manual for the Reasoning Type of Task Selection" (Hammer, 2016, p. 96), "learning support" included the feature of "differentiation": "Possibility of internal differentiation or natural differentiation, e.g., 'The learners have the option, depending on achieving a variety of difficult tasks'" (p. 96). Our study aims to address exactly this feature as a whole, but also to extend it with other specific features such as "openness". In contrast to Hammer, we consider tasks as a whole that can be worked on by all students at different levels, rather than a multitude of different tasks for different student levels.

Research Questions

Teachers face a complex and demanding job when considering the differentiation potential of tasks (e.g., when planning lessons). To adequately prepare teachers for this job during teacher training courses, more precise knowledge is needed about what teachers look at when assessing tasks in terms of their differentiation potential. How do teachers justify the selection or non-selection of specific tasks? Which task features do they focus on?

Previous research has paid little attention to the question of what competencies teachers have for recognizing tasks with differentiation potential. Therefore, we aimed to investigate teachers' capacity to reflect on this feature when specifically asked to do so. Consequently, we posed the following research questions:

1. To what extent do teachers succeed in correctly identifying the differentiation potential of tasks compared to experts?
2. Which task features do teachers refer to when assessing tasks in terms of their differentiation potential?
3. How can teachers' justifications for their assessments be categorized?

By investigating these questions, we intend to look deeper into teachers' reasoning processes during the selection and evaluation of tasks for differentiated instruction than has been done previously.

Method

The participants of our study were in-service teachers (N = 78) at secondary schools in Germany with several years of teaching experience (on average, more than nine years) who took part in an in-service teacher professional development course. A survey of teacher competencies related to the selection of appropriate adaptive tasks should reflect teachers' core activities as closely as possible. Therefore, we chose the scenario of lesson planning, which usually involves selecting suitable tasks for one's lessons from a large number of tasks in a textbook. The questionnaire, which was filled out by the teachers before the start of their professional development course, asked the participants to imagine a situation of planning a mathematics exercise phase with the goal to consolidate knowledge on comparing and adding fractions, with a heterogeneous group of students.

The questionnaire contained two types of tasks: tasks with differentiation potential according to the experts (i.e., suitable, see the example in Figure 2) and tasks that had low or no differentiation potential (i.e., unsuitable, see the example in Figure 6). The teachers had to provide some reasons for why they rated the tasks as "unsuitable," "rather unsuitable," "rather suitable," or "suitable." In addition, the questionnaire also contained mixed forms, which, for example, were only suitable for adaptive use to a limited extent. Such tasks showed only local characteristics of adaptive tasks (see the next section).

With this inductive qualitative method, we used the teachers' written rationales and assessments to identify which features teachers focus on when considering tasks in terms of adaptivity. In the questionnaire, we explicitly did not refer to any task feature to leave room for teachers to focus on task features that we may not have anticipated in advance.

Calculate.

a) $\frac{1}{4}+\frac{2}{3}$ b) $\frac{1}{3}+\frac{2}{5}$ c) $\frac{1}{2}+\frac{4}{9}$ d) $\frac{5}{6}-\frac{2}{5}$ e) $\frac{3}{4}-\frac{1}{3}$ f) $\frac{4}{5}-\frac{1}{2}$

How suitable do you think this task is for use in heterogeneous learning groups?

☐ unsuitable ☐ rather unsuitable ☐ rather suitable ☐ suitable

Describe aspects of the task that were relevant to you in making your decision.

Figure 6. Task 5 from the questionnaire. This is an example of a task that has almost no differentiation potential according to experts.

Developing the Questionnaire

The major challenge in developing the questionnaire was to vary the different tasks in terms of differentiation potential so that some tasks would clearly have such potential while others would not. To begin with, we were guided by the table in Figure 1 and systematically varied the deep structure, especially regarding "difficulty" and "openness." In addition, we varied the surface structure, and introduced, for example, parallel adaptive tasks and graded adaptive tasks, as the teachers were familiar with these formats.

It must be mentioned that, for example, the deep structure of parallel adaptive tasks does not necessarily mean that the surface structure of the tasks between the left- and right-hand sides must also differ in "difficulty." On the contrary, many textbooks showed that only the surface features (e.g., larger numbers) varied. In the end, we settled on eight tasks to sufficiently vary the questionnaire; moreover, the questionnaire could not be too extensive because we wanted the teachers to process everything in a satisfactory manner until the end, if possible.

To validate the task selection with regard to these requirements, we asked experts to evaluate and rate the tasks. The expert rating evaluated the tasks in terms of their differentiation potential and was conducted by three independent individuals (the authors), who have teaching experience and are involved in research and university teaching. In addition, the experts have been working on the topic of differentiation for several years. The agreement of the three experts was fair to good (with a Fleiss kappa value of $\kappa = .54$; see Table 2). The kappa statistic is useful for determining the degree of agreement between the ratings of multiple raters when the latter rate the same samples. The questionnaire was then designed in such a way that the tasks with varying differentiation potential (weak, medium, and strong) were randomly presented, and not marked to reflect their differentiation potential.

First, the teachers were asked to indicate, using a four-point Likert scale, how suitable each task was for heterogeneous learning groups (unsuitable, rather unsuitable, rather suitable, suitable; see Figures 2, 4 and 6). Next, the teachers were instructed to describe at least two features of each task that were relevant for their evaluation of the task. Second, we compared the teachers' evaluations of the tasks with the expert ratings. Finally, we analyzed teachers' comments using a specially developed rating manual (Table A in the Appendix).

Evaluation Method

A rating manual (see Table A in the Appendix) was developed to evaluate the teachers' open-ended responses. First, we developed the manual based on theoretical considerations, which also formed the basis of the questionnaire. To begin with, we considered the criteria of "openness" and "difficulty" and the adaptive task types, which differ in surface structure, as discussed in the Review-section. Furthermore, we focused on cognitive activities, when addressing student levels. A literature review that surveyed numerous task characteristics constituted an important basis for the features of our rating manual (see Figure 3 and Table B in the Appendix).

An analysis of teachers' reasons for their task assessments then expanded the number of characteristics that we researched in the literature in relation to classroom heterogeneity. Teachers mentioned a variety of features that, from our perspective, seemed to focus on other aspects than the differentiation potential of tasks. However, we were excited by these new features because they revealed what teachers were concerned with when selecting adaptive tasks.

During the evaluation, it became necessary to add and rewrite features with additional characteristics; as a result, the evaluation manual was continuously developed inductively. In this respect, our approach was guided by grounded theory (Glaser & Strauss, 1967). We started with the characteristics of "openness" and "difficulty" (both characteristics have been explained in detail earlier), referring to Jordan et al. (2006) and Wellenreuther (2004), who emphasized the "promotion of different forms of knowledge acquisition" and "grading according to the task difficulty," and to Bölsterli Bardy and Wilhelm (2018), who considered the "diversity of learning pathways" to be important.

The feature of "language" was mentioned both in the literature (see Figure 3; Anderson, 1996; Neubrand, 2002; Wellenreuther, 2004; Jordan et al., 2006; Blömeke et al., 2006; Maier et al., 2010; Leuders & Prediger, 2016) and by the teachers. This feature is relevant because it poses different challenges in terms of whether a task is designed to be linguistically complicated or simple.

In this context, "accessibility" (see Table A in the Appendix) as a task feature also turned out to be significant. "Accessibility" as a task feature means that a task contains introductory parts that facilitate successful entry, especially for low-achieving students (but also for all students). A students' successful entry into the processing of a task can occur, for example, through an explanation of the task, through simple subtasks, or through concrete examples. However, none of the articles in the literature source we examined described this feature (see Figure 3).

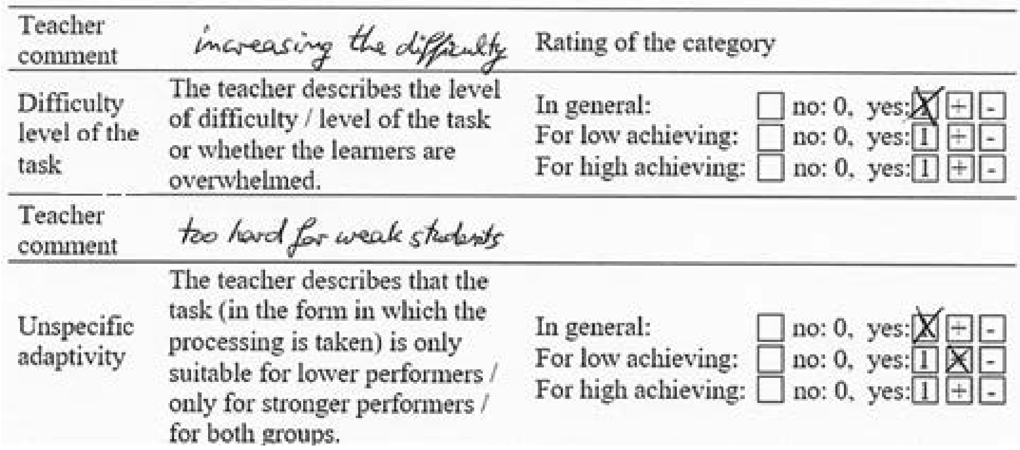
The feature "language" describes the linguistic requirements that students have to deal with in a task—for example, logical functions (negations, if-then connections) and universal or existential statements (an example task showing the importance of language can be found in Figure 5). Linguistic simplifications (as shown in the example tasks in Figure 1 and Figure 6) are especially useful for low-achieving students. A task can contain a description of how to proceed toward a possible solution (the task feature of "guidance") and/or additional hints that pre-structure the solution or help find a solution (the task feature of "support"). These two task features are especially helpful for low-achieving students.

These six features were the basis for the rating of the first completed questionnaire (answered by Teacher 1) and in the analysis of this questionnaire, these six features were supplemented by newly mentioned task features in the answers of Teacher 1. These additional features thus enlarged the rating scheme. This resulted in an extended category system, which was then applied to the next completed questionnaire (answered by Teacher 2), and so on.

Remarkably, some teachers' answers did not focus on differentiation potential but included other aspects that seemed important to the teachers. For example, teachers mentioned features such as "layout" or "task context" (see Table C and Table D in the Appendix). We did not want to simply summarize these mentions under one category; instead, we wanted to consider them in a more nuanced way because these were exactly the kind of answers that were interesting to us and that could be very helpful when it comes to planning teacher training. Therefore, the complete category system finally comprised 22 features (see Table A in the Appendix). Since 78 questionnaires were included, this result indicates a certain saturation of categories within this sample.

We rated each feature based on whether it was mentioned in relation to low- or high-achieving students. We used the rating manual to evaluate teacher comments in the questionnaire. After rater training on the 22 features and an intensive introduction to the rating manual, the ratings of six further raters yielded a Fleiss kappa value of κ = .65 (substantial agreement). One member of the expert team then applied the rating scheme to all 78 questionnaires. The concrete procedure was as follows:

If a category matched a teacher's comment, the category was assigned the value "1"; "–" if the comment was negative and "+" if the comment was positive. If the commentary was formulated with respect to high-achieving or low-achieving students, the commentary was assigned the value "1" in the evaluative category of "generally." For average-, high-, and low-achieving students, the values "1", "+", and "–" were assigned, respectively. Two rating examples can be found in Figure 7.



Teacher comment	increasing the difficulty	Rating of the category
Difficulty level of the task	The teacher describes the level of difficulty / level of the task or whether the learners are overwhelmed.	In general: ☐ no: 0, yes: ☒ ☐+ ☐- For low achieving: ☐ no: 0, yes: ☐1 ☐+ ☐- For high achieving: ☐ no: 0, yes: ☐1 ☐+ ☐-
Teacher comment	too hard for weak students	
Unspecific adaptivity	The teacher describes that the task (in the form in which the processing is taken) is only suitable for lower performers / only for stronger performers / for both groups.	In general: ☐ no: 0, yes: ☒ ☐+ ☐- For low achieving: ☐ no: 0, yes: ☐1 ☒ ☐- For high achieving: ☐ no: 0, yes: ☐1 ☐+ ☐-

Figure 7. Two rating examples.

Participating Teachers

The participants of the study were in-service teachers (N = 78) at secondary schools in Germany with several years of teaching experience (see Table 1). The teachers participated voluntarily and were motivated because they had registered for a teacher training course. The group consisted of 46 female and 32 male teachers.

Table 1: Number of teachers and their teaching experience.

Teaching experience	Total
Up to 3 years	3
Between 3 and 9 years	13
More than 9 years	25
No information	37

Results

To answer the first research question, we compared the participants' evaluations with expert evaluations (see Table 2).

Table 2: Teacher (N=78) judgements compared to expert (N=3) judgements.

Rating	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
Mean teacher judgment	2.53	2.19	2.54	2.97	2.62	2.47	2.42	2.82
SD teacher	.729	.873	.903	.974	.856	.959	.956	1.010

The comparison of the ratings between the teachers and the experts yielded interesting results. While the experts exhausted the scale using extremes, the teachers tended to rely on the middle of the response scale (this is also shown by the moderate standard deviations). The experts clearly rated Tasks 0, 2, 3, 4, and 7 as suitable and Tasks 1, 5, and 6 as unsuitable (for Task 0, see Figure 2; Task 3 is shown in Figure 4, and Task 5 in Figure 6).

The teachers assigned the highest values to Tasks 3 (2.97) and 7 (2.82), which were followed by Tasks 2 (2.54) and 4 (2.62). This corresponds to the middle range of the scale, which means that the teachers did not agree in their evaluations here. Interestingly, the experts rated Task 0 as suitable (mean teacher judgement 2.53; what means rather unsuitable/rather suitable) and Task 5 as unsuitable (mean teacher judgement 2.47; what means rather unsuitable/rather suitable).

The teachers were first asked to assess the differentiation potential of a task and then to provide reasons for their assessment. For this purpose, we looked at one task that, in our view, had high differentiation potential (Task 0, see Figure 2) and at another task that had low or no differentiation potential (Task 5).

To start, we rated the teachers' open-ended responses using our coding manual. Then, we sorted these codes according to how the teachers rated a task as a whole. Tables C and D in the Appendix present an overview of this process. All features were mentioned across all evaluation categories (from unsuitable to suitable). Consequently, a more differentiated view of the individual statements within the categories is required—that is, the naming of a feature alone is not sufficient to understand how the teachers assessed the differentiation potential of a task.

The excerpts in Table 3 provide some examples of the teachers' comments in the questionnaire on Task 0. This is a selection of justifications from the first twenty-five questionnaires. The excerpts clearly show how differently the same feature was evaluated and how differentiated or undifferentiated a feature was considered to be. It is interesting to note the different kinds of argumentation used to justify why someone considered a task to be suitable or unsuitable. While those who considered a task suitable tended to look at the task features that we have described, those who considered a task rather unsuitable listed features that were irrelevant in terms of differentiation potential, at least from our perspective. In addition, those who recognized the differentiation potential of a task tended to describe the task's individual features in more detail than those who did not recognize its differentiation potential.

Table 3: A selection of teachers' justifications for their rating decisions on Task 0.

Assessment of Task 0				
Teacher Rating	Unsuitable 1% (1)	Rather unsuitable 32% (25)	Rather suitable 55% (43)	Suitable 12% (9)
Task features				
Difficulty				
Comments without reference to adaptivity	*Not obvious which task is difficult or easy.*	*Too difficult for low-achieving students.* *Little challenge for strong students.* *"Only" reproduction of multiplication and division.* *Only difficult numbers.*	*Level of difficulty.* *Difficulty level very high.* *Too difficult for low-achieving students.* *Differentiation through blocks and larger numbers.* *Too few tasks at the basic level.* *Difficult due to ambiguous tasks.*	*Simple calculation paths are possible for all students.*
Comments with reference to adaptivity	-------	*Left-hand side is easier due to small numbers.* *The structure of levels is non-linear [Subtask No. 2 more difficult than No. 3].*	*Increase in difficulty.* *Automatically more difficult/easier.* *Easy entry.* *Working toward the same goal at different levels.* *Differentiation is evident not in the number of tasks but in the difficulty.* *Different levels without this being immediately visually recognizable.* *From left to right somewhat more complicated (larger numbers, more computational effort).* *From Task 1 to 4 more complex strategies become necessary.*	*Task blocks are graded in difficulty [Subtask No. 2 left: working backwards, Subtask No. 4 left: puzzling].* *Easy / hard tasks.* *Increasing difficulty.* *Easy start.* *Starts easy and gets harder.* *Increasing difficulty.* *Level increase can be seen.* *Left-hand side less complicated than right-hand side.* *Number of possible solutions increases.*
Openness				
	Open solution processing.	*Too open for low-achieving students.* *Results are unambiguous.* *Hardly any different calculation methods are possible.*	*Multiple solution schemes are necessary.* *Different strategies can be used.* *Openness of subtasks.* *Tasks with multiple solutions are available.*	*Different approaches.* *Open and closed tasks.*

Note: Each sentence was a response in the questionnaire by a different teacher. Recall also that all three experts rated this task as suitable in terms of its differentiation potential.

Table 4: A selection of teachers' justifications for their rating decisions on Task 5.

Assessment of Task 5				
Teacher Rating	Unsuitable 27% (20)	Rather unsuitable 49% (36)	Rather suitable 19% (14)	Suitable 5% (4)
Task features				
Difficulty				
Comments without reference to adaptivity	*Too simple.*	*A low-achieving student only gets as far as c).*	*The beginning of the task would have to be even easier to rate the task as "suitable".* *Increasing difficulty.*	------
Comments with reference to adaptivity	*All students have to calculate the same tasks.* *No differentiation via task level.* *All tasks with similar requirements.* *High-achieving students completed all tasks quickly.* *No different level of difficulty.*	*No variation in difficulty.* *No increased difficulty.* *Equal difficulty in subtasks.* *Simple structureless sequence of tasks.* *Not a challenge for high-achieving students.* *Exercise only for low-achieving students.*	*Only few level differences between the tasks.* *All tasks are similarly difficult.*	*A standard task for low-achieving students.*
Openness				
	"Working out" the algorithm with a unique solution.	*All students do the same.* *All students process the same tasks.*	-------	------

Note: Four teachers did not provide an assessment of this task. Each sentence was a response in the questionnaire by a different teacher. Recall also that all three experts rated this task as unsuitable in terms of its differentiation potential.

In our study, we distinguished between statements that explicitly referred to adaptivity and those that focused on other task features. For example, statements such as "too difficult for low-achieving students" (see Table 3, assessment of Task 0) focused on low-achieving students but did not look at the difference of difficulty for high-achieving students. In contrast, a statement such as "working toward the same goal at different levels" focused on a task's diversity of requirements.

It was also important to us that teachers not only recognize the differentiating potential of tasks but also identify the tasks that do not have this potential. This is particularly important because there are numerous tasks in textbooks that do not have any differentiation potential. In the answers related to Task 5 (see Table 4), it was clear that there were differences among the teachers regarding their assessment of the task. Statements that did not contain any reference to adaptivity included, for example, descriptions like "too simple". The teachers who, for example, made statements such as "no different level of difficulty" were more likely to arrive at the assessment "unsuitable" as far as the differentiation potential of the task was concerned.

Regarding Task 5, some teachers justify the suitability of this task as "a standard task for low-achieving students". It is interesting that these teachers seem to equate differentiation with task accessibility for low-achieving students (see Table 4).

As the responses in Tables 3 and 4 show, the teachers focused on different features, which we then included in our rating manual. In total, 22 features emerged. For a complete illustration, see Table A in the Appendix. The aim of this category system was to work out the task features that the teachers looked at. Therefore, the same features could receive both an approving and a disapproving evaluation. For example, "difficulty" could mean that the task as a whole was too difficult. At the same time, it could also describe the different requirements behind the individual subtasks. Consequently, it is important to look at individual statements regarding task features in a nuanced way, as was done in Tables 3 and 4. In this section, we have taken a deeper look at teachers' different assessments of tasks in relation to various features.

	Start- and added features	**Sum across all tasks**	**All features sorted by frequency**	**Sum across all tasks**
Features based on theoretical considerations (Start-features)	Openness	134	Routine/drill	256
	Difficulty	200	Difficulty	200
	Language	87	Layout	191
	Guidance	26	Openness	134
	Support	56	Goal differentiation	101
	Accessibility	52	Language	87
Features additionally mentioned by teachers (Added features)	Routine/drill	256	Context/application	73
	Layout	191	Motivation	64
	Goal differentiation	101	Problem solving	62
	Context/application	73	Support	56
	Motivation	64	Transparent structure	56
	Problem solving	62	Accessibility	52
	Transparent structure	56	Size	49
	Size	49	Creativity	41
	Creativity	41	Prerequisites	27
	Prerequisites	27	Guidance	26
	Self-checking	18	Self-checking	18
	Presentation	18	Presentation	18
	Change of mathematical representation	16	Change of mathematical representation	16
	Reasoning	8	Reasoning	8
	Generalizing	3	Generalizing	3

Figure 8. Sum of task features mentioned by teachers across all eight tasks from the questionnaire (Features based on theoretical considerations highlighted in gray).

In the following section, we will discuss the most frequently mentioned features. Such features can be divided into those that are close to our interests and those that are rather irrelevant from our point of view but seem to be highly important for the teachers. The most frequently mentioned characteristics were “routine/drill,” “difficulty,” “layout,” “openness,” and “goal differentiation” (see Figure 8 and Table A in the Appendix). Many other more frequently focused features supplemented the six features with which we started the analyses of teacher comments (see Figure 8).

What do teachers mean when they mention or write down the task feature of “routine/drill”? It mainly had to do with procedural knowledge, which focuses on automation and calculation. It was also associated with rather low cognitive activation. The focus here was on the practice of calculations. We have already explained the task feature “difficulty” in more detail in the Section “The Deep Structure of Tasks”. Difficulty was the second most frequently mentioned feature.

The feature of “layout” was mentioned remarkably frequently. This feature referred to the reading direction in the tasks, such as in Task 0 (see Figure 2), where the arithmetic operation had to be performed vertically. Also, the parallel arrangement of tasks by a vertical line in the middle elicited the mention of the “layout” feature.

We have discussed the task feature of “openness” in detail with examples using Tasks 0 (see Table 3) and 5 (see Table 4). In this case, the different evaluations were evident as well: on the one hand, “openness” was seen as conducive to working in heterogeneous groups because it could ensure adaptivity. On the other hand, some teachers saw this feature as being too much of a challenge or as being too unclear for their students.

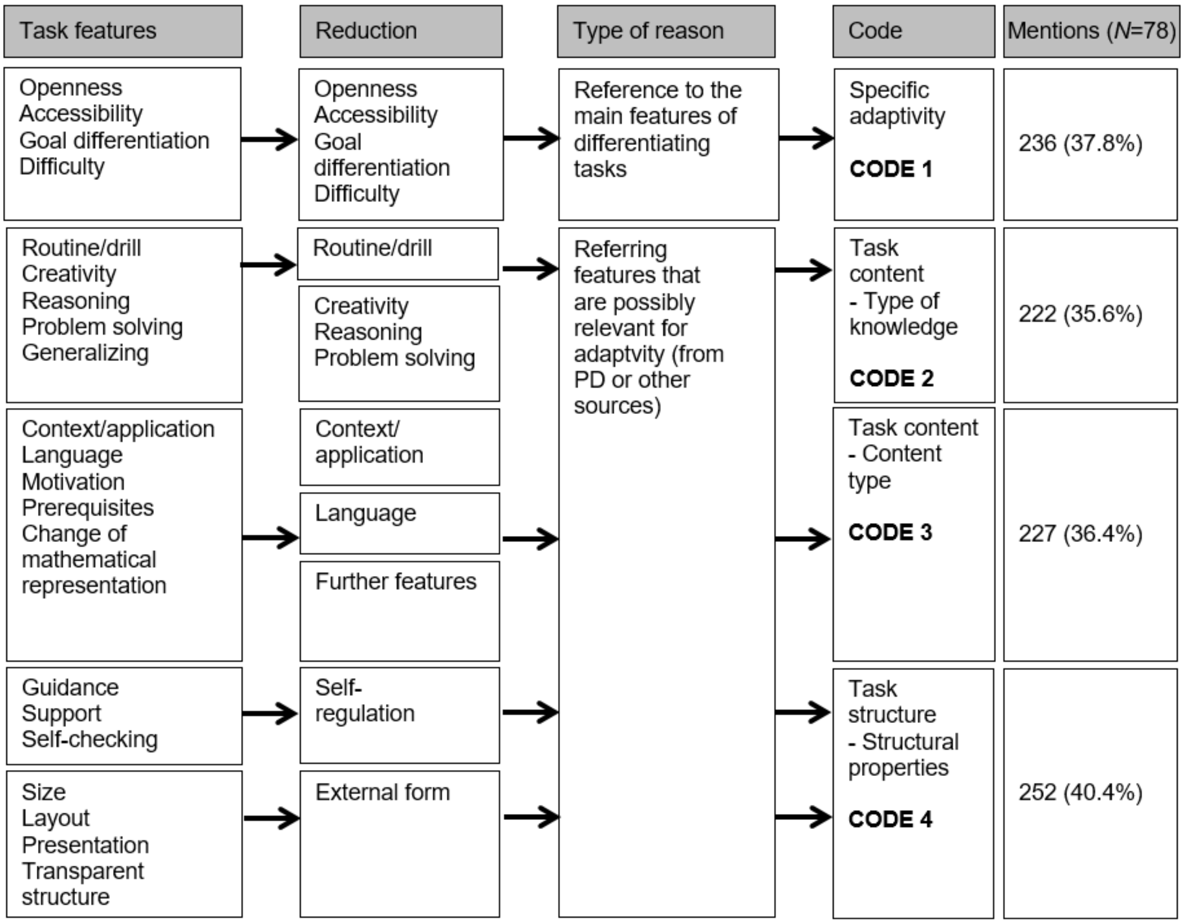


Figure 9. Condensed category system (see Holzäpfel et al., 2019, p. 372).

We have not yet discussed the feature of "goal differentiation" in detail. This feature involves setting different objectives for different students. Consequently, students also receive different tasks, with different degrees of difficulty. Basically, this feature is closely related to the openness and difficulty features but has a more formal character, especially with regard to the evaluation of student performance.

Based on these nuanced analyses, we then developed a condensed category system that summarized the core aspects of adaptive tasks. We proceeded in several steps, which are illustrated in Figure 9. To cluster the teachers' answers (Research Question 3), we condensed 21 existing features (without the feature "miscellaneous"; see Table A in the Appendix). We carried out this procedure because not all codes could be mentioned for each of the eight tasks. When it comes to cluster analysis, it is necessary to name the individual codes for each task to ensure the comparability of the codes. Figure 9 displays the results. For example, a teacher may mention a task feature eight times across all eight tasks, although this feature is actually characteristic of only two tasks. Thus, we had to develop a categorization that would consider the actual occurrence of the features across all tasks. The result is the grid on the right-hand side of Figure 9 (Code 1 to 4). Overall, teachers in the questionnaire mention the four codes in a balanced way in their justifications. There is a slight focus on Code 4 "Task structure," which has a share of about 40%. The low percentage (38%) of justifications with Code 1 "Specific adaptivity" shows that most teachers did not focus on the central features of adaptive tasks in their justifications.

Discussion

In this study, we examined how teachers assess the differentiation potential of tasks compared to expert assessment. First, we analysed which characteristics could be used to identify the differentiation potential of tasks. An analysis of the teachers' answers in the questionnaire showed that teachers' assessments differed from the assessments of experts. By looking at the patterns of reasoning in a nuanced way, we were able to gain insights into the different features that the teachers considered (e.g., Tables 3 and 4).

The first research question examined how teachers recognize the differentiation potential of tasks compared to experts. In this case, the questionnaire revealed clear differences in assessment. However, standard deviations also showed that there were different perspectives on the differentiation potential of tasks among the teachers and that there were certain subgroups among the teachers who assessed this issue more in the direction of the expert opinion or perceived it differently to the experts. Cautious assessments of the tasks were noticeable (the teachers did not assign extreme values when assessing whether the tasks had differentiation potential or not); this could indicate uncertainty on behalf of the teachers (see Tables 3 and 4). Alternatively, it could also represent an implicit acknowledgment from the teachers that the manner in which the task is implemented in the classroom impacts its differentiation potential.

The reasons that the teachers provided to justify their statements were addressed by our second research question. We were interested in finding out which task features the teachers relied on when judging tasks in terms of their differentiation potential. In our analyses, we learned that the teachers had different perspectives on this matter. Some teachers expressed views that were close to expert opinion, while others looked at unexpected features, such as "layout," or "routine/drill" (see Figure 8). Likewise, there was a strong orientation toward procedures for calculation, which many teachers saw as being connected with the question of adaptivity. Task openings that led to different levels of processing—including, for example, "problem solving," the use of other "mathematical representations," or "creativity" in processing—were only partially recognizable here, perhaps implying that a wider understanding of "openness" should be adopted in teacher training.

The third research question aimed to categorize the task features mentioned by the teachers. To accomplish this goal, we started with theoretical considerations (see Table B in the Appendix) and then added further categories based on the teachers' responses. Interestingly, the teachers mentioned numerous features that were not necessarily associated with adaptivity. However, it is important to consider this when planning a teacher-training program. Through our study, teachers' current perspectives on adaptive tasks became clear, and we used the latter as important starting points for examining the core aspects of adaptive tasks.

We found that the teachers who had a different opinion from the experts were more likely to evaluate the surface structure of tasks. On the one hand, the teachers perceived certain features as confusing because, for example, the task types and layouts were unfamiliar or the mathematical activities were uncommon (e.g., Task 0 in Figure 2). On the other hand, the teachers focused quite strongly on mathematical procedures, which were especially calculation oriented. In this context, differences were more likely to arise at the level of complication—for example, when the numbers are increased, which did not necessarily correspond to a different level of mathematical demand.

Methodologically, we faced numerous challenges in this study. We wanted to design the eight question items as close to the textbooks (used in practice phases) as possible to make the situation as authentic as possible for the teachers. However, we had to make compromises as the participating teachers worked in schools with different textbooks. Another challenge involved the fact that our questionnaire was limited to a small sample of tasks. Clearly, the findings would have been more nuanced had we added more task types and other content areas.

Another difficulty had to do with the fact that the tasks in the questionnaire did not necessarily match the teachers' classes. This meant that the teachers had to imagine a fictitious class for which they were selecting the tasks. It also became clear from the teachers' feedback that the surface structure of some tasks from the questionnaire are unknown to their own students and therefore there was already a problem in using these tasks in their own lessons. This might have served as an additional reason for some teachers to judge these particular tasks as unsuitable with regard to their suitability for differentiation in the classroom.

From a teacher perspective, our results are interesting in several respects. The task features we grouped in Figure 9 under the Codes 1 to 4 can be used concretely to get a focused view on tasks when planning lessons. If a teacher has discovered or developed concrete adaptive tasks in his or her lesson planning (for example, using the surface structure or selected features from Code 1), this does not mean that his or her students can automatically work with such tasks. A process of instruction is needed to familiarize students with adaptive tasks (see the tasks in Figure 2, Figure 4 or Figure 5). At which subtask number should a high-achieving student begin? A low-achieving student can start with the first subtask, and the teacher does not expect this student to work on all subtasks until the last one. A high-achieving student might be encouraged to work only on the last subtasks. In a parallel adaptive task, a student is allowed to choose one task column and they know that the left side is generally more appropriate for low-achieving students.

All in all, we believe our analysis was highly informative and provided important insights into the planning of further professional development courses for teachers. More specifically, the deep structure of tasks seems to be an area in which teachers need to develop further expertise; with the implication that it should be more of a focus of teacher training courses. However, teachers' perspectives must also be taken into account and critically reflected upon; when it comes to differentiation potential, the teachers seemed to consider various task features very differently than did the task-design experts.

Acknowledgements

We would like to thank the German Centre for Teacher Education in Mathematics (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.

References

Anderson, J. (1996). Some teachers' beliefs and perceptions of problem solving. In P. 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.

Barzel, B., Leuders, T., Prediger, S., & Hußmann, S. (2013). Designing tasks for engaging students in active knowledge organization. In C. Margolinas, A. Watson, M. Ohtani, J. Ainley, J. Bolite Frant, M. Doorman, C. Kieran, A. Leung, P. Sullivan, D. Thompson, & Y. Yang (Eds.), *ICMI Study 22 on Task Design:Pproceedings of the study conference* (pp. 285–294). ICMI.

Beck, E., Baer, M., Guldimann, T., Bischoff, S., Brühwiler, C., Müller, P., Niedermann, R., Rogalla, M., & Vogt, F. (2008). *Adaptive lehrkompetenz* [Adaptive teaching skills]. Waxmann.

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 [Analysis of the quality of tasks from a didactic and subject perspective: A general model and its exemplary implementation in the School subject Mathematics]. *Unterrichtswissenschaft, 34*(4), 330–357.

Bölsterli Bardy, K., & Wilhelm, M. (2018). Von kompetenzorientierten zu kompetenzfördernden Aufgaben im Schulbuch [From competence-oriented to competence-promoting tasks in the textbook]. *Erziehung und Unterricht*, *168,* 1/2, 121–129.

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education*, *40*(2), 119–156.

Bromme, R., Seeger, F., & Steinbring, H. (1990). *Aufgaben als Anforderungen an Lehrer und Schüler* [Tasks as demands on teachers and students]. Aulis.

Brown, A. L. (1984). Metakognition, Handlungskontrolle, Selbststeuerung und andere, noch geheimnisvolle Mechanismen [Metacognition, action control, self-control, and other still mysterious mechanisms]. In F. E. Weinert & R. H. Kluwe (Eds.), *Metakognition, Motivation und Lernen* (pp. 60–109). Kohlhammer.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively guided instruction*. Heinemann.

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

Clarke, D. M., Cheeseman, J., Roche, A., & Van Der Schans, S. (2014). Teaching strategies for building student persistence on challenging tasks: Insights emerging from two approaches to teacher professional learning. *Mathematics Teacher Education and Development, 16*(2), 46–70.

Doyle, W. (1983). Academic work. *Review of Educational Research, 53*(2), 159–199. https://doi.org/10.3102%2F00346543053002159

Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist, 23*(2), 167–180. https://doi.org/10.1207/s15326985ep2302_6

Glaser, B. G., & Strauss, A. L. (1967). *A discovery of grounded theory*. Aldine Publishing.

Hammer, S. (2016). *Professionelle Kompetenz von Mathematiklehrkräften im Umgang mit Aufgaben in der Unterrichtsplanung: theoretische Grundlegung und empirische Untersuchung* [Professional competence of mathematics teachers in dealing with tasks in lesson planning: theoretical grounding and empirical investigation; Doctoral dissertation, Ludwig–Maximilians–Universität München]. https://edoc.ub.uni-muenchen.de/20439/1/Hammer_Sabine.pdf

Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* [Teaching quality and teacher professionalism: diagnosis, evaluation and improvement of teaching]. Klett-Kallmeyer.

Hengartner, E., Hirt, U., Wälti, B., & Primarschulteam Lupsingen (2006). *Lernumgebungen für Rechenschwache bis Hochbegabte: Natürliche Differenzierung im Mathematikunterricht* [Learning environments for low- to high-achieving students in arithmetic: Natural differentiation in mathematics education]. Klett & Balmer.

Heymann, H. W. (1991). Innere Differenzierung im Mathematikunterricht [Internal differentiation in mathematics education]. *Mathematik lehren, 49*, 63–66.

Hiebert, J., & Wearne, D. (1997). Instructional tasks, classroom discourse and student learning in second grade arithmetic. *American Educational Research Journal, 30*(2), 393–425. https://doi.org/10.3102/00028312030002393

Hirt, U., & Wälti, B. (2008). *Lernumgebungen im Mathematikunterricht: Natürliche Differenzierung für Rechenschwache und Hochbegabte* [Learning environments in mathematics education: Natural differentiation for low- and high-achieving students]. Klett-Kallmeyer.

Holzäpfel, L., Lacher, M., Leuders, T., & Rott, B. (2018). *Problemlösen lehren lernen: Wege zum mathematischen Denken* [Learning to teach problem solving: pathways to mathematical reasoning]. Klett-Kallmeyer.

Holzäpfel, L., Leuders, T., & Bardy, T. (2019). Preparing in-service teachers for the differentiated classroom. In M. Graven, H. Venkat, A. Essien, & P. Vale (Eds.), *Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education*, *2*, 60–109. PME.

Hußmann, S. & Prediger, S. (2007). Mit Unterschieden rechnen - Differenzieren und Individualisieren [Reckoning with differences: Differentiating and Individualizing]. *Praxis der Mathematik in der Schule, 49*(17), 2–8.

Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., 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. https://pure.mpg.de/rest/items/item_2100753_3/component/file_2197661/content

Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., & Brunner, M. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht [Tasks in the COACTIV project: testimonies of cognitive activation potential in German mathematics education]. *Journal für Mathematik-Didaktik, 29*(2), 83–107. https://doi.org/10.1007/BF03339055

Krauthausen, G., & Scherer, P. (2010). *Umgang mit Heterogenität: Natürliche Differenzierung im Mathematikunterricht der Grundschule* [Dealing with heterogeneity: Natural differentiation in elementary school mathematics classrooms]. IPN. http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/ Handreichung_Krauthausen-Scherer.pdf

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project.* Springer. https://doi.org/10.1007/978-1-4614-5149-5

Leuders, T. (2015). Aufgaben in Forschung und Praxis [Tasks in research and practice]. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Eds.), *Handbuch Mathematikdidaktik* [Handbook mathematics education] (pp. 433-458). Springer.

Leuders, T., & Holzäpfel, L. (2011). Kognitive Aktivierung im Mathematikunterricht [Cognitive activation in mathematics education]. *Unterrichtswissenschaft, 39*, 213–230.

Leuders, T., & Prediger, S. (2016). *Flexibel differenzieren und fokussiert fördern im Mathematikunterricht* [Flexible differentiation and focused support in mathematics teaching]. Cornelsen Scriptor.

Lipowsky, F. & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *Ricercazione, 7*(2), 27–51.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423–458. https://doi.org/10.2307/1170650

Maier, U., Kleinknecht, M., Metz, K., & Bohl, T. (2010). Ein allgemeindidaktisches Kategoriensystem zur Analyse des kognitiven Potenzials von Aufgaben [A general didactic category system for analyzing the cognitive potential of tasks]. *Beiträge zur Lehrerinnen- und Lehrerbildung, 28*(1), 84–96.

McDougall, D. E. (2004). *School mathematics improvement: Leadership handbook*. Thomson Nelson.

Müller, G., & Wittmann, E. Ch. (1998). *Das Zahlenbuch: Lehrerband* [The numbers book: Teacher's volume]. Klett.

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: Self-acting in student work phases in lessons of the TIMSS video study; Doctoral dissertation, Freie Universität Berlin].

Pozas, M., Letzel, V., & Schneider, C. (2020). Teachers and differentiated instruction: Exploring differentiation practices to address student diversity. *Journal of Research in Special Educational Needs, 20*(3), 217–230. https://doi.org/10.1111/1471-3802.12481

Prediger, S., & Scherres, C. (2012). Niveauangemessenheit von Arbeitsprozessen in selbstdifferenzierenden Lernumgebungen: Qualitative Fallstudie am Beispiel der Suche aller Würfelnetze [Level appropriateness of work processes in self-differentiating learning environments: Qualitative case study using the example of searching all cube nets]. *Journal für Mathematik-Didaktik, 33*, 143–173. https://doi.org/10.1007/s13138-012-0035-9

Prediger, S., Roesken-Winter, B., & Leuders, T. (2019). Which research can support PD facilitators? Strategies for content-related PD research in the Three-Tetrahedron Model. *Journal of Mathematics Teacher Education, 22*, 407–425. https://doi.org/10.1007/s10857-019-09434-3

Prediger, S., Barzel, B., Hußmann, S., & Leuders, T. (2021). Towards a research base for textbooks as teacher support: The case of engaging students in active knowledge organization in the KOSIMA project. *ZDM Mathematics Education*. [Online first]. https://doi.org/10.1007/s11858-021-01245-2

Russo, J., Minas, M., Hewish, T., & McCosh, J. (2020). Using prompts to empower learners: Exploring primary students' attitudes towards enabling prompts when learning mathematics through problem solving. *Mathematics Teacher Education and Development, 22*(1), 48–67. https://mted.merga.net.au/index.php/mted/article/view/592

Scherres, C. (2012). *Niveauangemessenes mathematisches Arbeiten in einer selbstdifferenzierenden Lernumgebung. Eine qualitative Fallstudie am Beispiel einer Würfelnetz-Lernumgebung* [Level-appropriate mathematical work in a self-differentiating learning environment: A qualitative case study using the example of a cube net learning environment; Doctoral dissertation, IEEM Dortmund]. https://doi.org/10.1007/978-3-658-02083-5

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist, 23*(2), 145–166.

Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer"-Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht ["Observer"-validation of a video-based instrument to capture professional perceptions of teaching]. *Zeitschrift für Pädagogik, 56*, 296–306.

Snow, R. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 13–60). W. H. Freeman.

Stein, M., 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.

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.

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., Mousley, J., & Jorgensen, R. (2009). Tasks and pedagogies that facilitate mathematical problem solving. In B. Kaur, Y. B. Har, & M. Kapur (Eds.), *Mathematical Problem Solving: Yearbook 2009* (pp. 17–42). World Scientific Publishing Co. https://doi.org/10.1142/9789814277228_0002

Sullivan, P., Borcek, C., Walker, N., & Rennie, M. (2016). Exploring a structure for mathematics lessons that initiate learning by activating cognition on challenging tasks. *The Journal of Mathematical Behavior, 41*, 159–170. https://doi.org/10.1016/j.jmathb.2015.12.002

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.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
Watson, A., & Mason, J. (2006). Seeing exercise as a single mathematical object: Using variation to structure sense-making. *Mathematical Teaching and Learning, 8*(2), 91–111. https://doi.org/10.1207/s15327833mtl0802_1
Wellenreuther, M. (2004). *Lehren und Lernen – aber wie? Empirisch-experimentelle Forschung zum Lehren und Lernen im Unterricht* [Teaching and learning - but how? Empirical-experimental research on teaching and learning in the classroom]. Schneider Hohengehren.
Wittmann, E. C., & Müller, G. N. (1990). *Handbuch produktiver Rechenübungen* [Handbook of productive arithmetic exercises]. Band II. Klett.
Zaslavsky, O., & Sullivan, P. (Eds.). (2011). *Constructing knowledge for teaching secondary mathematics*. Springer. https://doi.org/10.1007/978-0-387-09812-8

Authors

Thomas Bardy
University of Education Freiburg
Kunzenweg 21, D-79117 Freiburg im Breisgau, Germany

Lars Holzäpfel
University of Education Freiburg
Kunzenweg 21, D-79117 Freiburg im Breisgau, Germany
email: lars.holzaepfel@ph-freiburg.de

Timo Leuders
University of Education Freiburg
Kunzenweg 21, D-79117 Freiburg im Breisgau, Germany

Appendix

Table A: Rating manual.

Feature	Description
Openness	The teacher describes the openness of the (sub-)task, which allows for different / individual solutions / approaches or results (problem solving required).
Difficulty	The teacher describes the level of difficulty / level of the task or whether the learners are overwhelmed.
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.
Guidance	The teacher states that the task provides guidelines (e.g., possible outcomes) or a pattern for approaching the work.
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).
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).
Empirically supplemented	
Goal differentiation	The teacher states that the task pursues different content goals for high-achieving and low-achieving students.
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").
Size	The teacher describes the scope / processing time of the task, the amount of tasks, or that many types of content are present in the task.
Context/ application	The teacher states that the task contains an application reference, illustrative content, (everyday) context, or is abstract.
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.
Motivation	The teacher states that the learners are motivated by the task, that the task is varied, or that frustration tolerance is high.
Self-checking	The teacher states that the task is designed to encourage learners' independence.
Presentation	The teacher states that the task contains representations or content that is unusual and potentially confusing to the learner.
Transparent structure	The teacher states that the approach to working on the (sub-)task(s) is (completely) clear, confusing, or complex to the learner.
Creativity	The teacher states that the task stimulates learners' creativity (own ideas, concepts, objects, ...).
Reasoning	The teacher states that the task requires an explanation of / reflection on a procedure or an argumentation from 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.
Generalizing	The teacher states that the task requires learners to recognize patterns and to generate or generalize conjectures or questions.

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.
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.
Miscellaneous	For example, the task is well suited for partner work; independence is required.

Table B: Task features with representative examples and literature references.

Task features	**At least one mention of the feature (*N* = 78 teachers)**	**Representative examples from teachers' justifications in the questionnaire**	**Task features in the literature**
Openness	.72	*Different approaches with multiple solutions.*	Anderson, 1996; Neubrand, 2002; McDougall, 2004; Jordan et al., 2006; Sullivan, 1999; Maier et al., 2010; Leuders & Prediger, 2016; Bölsterli Bardy & Wilhelm, 2018
Difficulty	.88	*The task starts easy and becomes more difficult.*	Stein & Smith, 1998; Sullivan, 1999; Wellenreuther, 2004; Blömeke et al., 2006; Leuders & Prediger, 2016; Hammer, 2016
Language	.69	*Language difficulties among pupils do not carry any weight.*	Anderson, 1996; Wellenreuther, 2004; Jordan et al., 2006; Maier et al., 2010; Leuders & Prediger, 2016
Guidance	.27	*Confusion among students due to missing arithmetic symbols.*	Stein et al., 1996; Stein & Smith, 1998; Jordan et al., 2006; Hammer, 2016; Bölsterli Bardy & Wilhelm, 2018
Support	.51	*Great need for explanation for low-achieving students.*	Stein et al., 1996; Stein & Smith, 1998; Sullivan, 1999; Jordan et al., 2006; Hammer, 2016; Bölsterli Bardy & Wilhelm, 2018
Accessibility	.40	*Easy start with the task.*	Leuders & Prediger, 2016; Sullivan, 1999
Empirically supplemented			
Goal differentiation	.65	*Working toward the same goal at different levels.*	
Routine/drill	.92	*Schematic processing of package tasks.*	Stein & Smith, 1998; Neubrand, 2002; Jordan et al., 2006; Hammer, 2016
Size	.38	*Uniform and adjusted scope.*	Anderson, 1996; Stein et al., 1996; Jordan et al., 2006
Context/application	.63	*An everyday reference and explicit actions are missing.*	Sullivan, 1999; Neubrand, 2002; McDougall, 2004; Blömeke et al., 2006; Maier et al., 2010; Bölsterli Bardy & Wilhelm, 2018
Layout	.83	*Structuring not clear (How do I proceed with the task?).*	Neubrand, 2002; Jordan et al., 2006; Maier et al., 2010; Leuders & Prediger, 2016; Hammer, 2016; Bölsterli Bardy & Wilhelm, 2018

Motivation	.47	*Existing reverse task provides motivation for the stronger students.*	Sullivan, 1999
Self-Checking	.21	*No self-control possible.*	Stein & Smith, 1998
Presentation	.23	*The required calculation method for the task is confusing.*	
Transparent structure	.47	*Clear structure of the task.*	Wellenreuther, 2004; Jordan et al., 2006; Maier et al., 2010; Hammer, 2016; Bölsterli Bardy & Wilhelm, 2018
Creativity	.37	*Different strategies can be used in the task.*	Anderson, 1996; Stein & Smith, 1998; Sullivan, 1999; Neubrand, 2002; Maier et al., 2010; Blömeke et al., 2006; Leuders & Holzäpfel, 2011; Hammer, 2016; Leuders & Prediger, 2016; Bölsterli Bardy & Wilhelm, 2018
Reasoning	.08	*The task is only "rather suitable" because no justifications are required.*	Stein et al., 1996; Stein & Smith, 1998; McDougall, 2004; Jordan et al., 2006
Problem solving	.51	*Different problem solving skills are required.*	Sullivan, 1999; Blömeke et al., 2006; Jordan et al., 2006; Holzäpfel et al., 2018
Generalizing	.04	*The ability to generalize is necessary.*	
Prerequisites	.26	*The task does not address different levels of prior knowledge.*	Stein & Smith, 1998; Leuders & Prediger, 2016
Change of mathematical representation	.18	*The task does not use different mathematical forms of representation.*	

Table C: Assessment of Task 0.

Assessment of Task 0				
Number of mentions of task features				
All assessments	**Unsuitable**	**Rather unsuitable**	**Rather suitable**	**Suitable**
Difficulty(47) Layout(38) Routine/drill(34) Problem solving(23) Openness(19) Goal diff.(19) Accessibility(18) Transparent Structure(16) Motivation(15) Presentation(14) Self-checking(13) Creativity(12) Guidance(8) Support(7) Size(5) Prerequisites(4) Language(4) Context/application(3) Reasoning(2)	Guidance(1) Difficulty(1) Layout(1) Transparent structure(1)	Layout(14) Difficulty(13) Routine/drill(12) Openness(9) Presentation(7) Goal differentiation(7) Problem solving(6) Creativity(6) Transparent structure(6) Support(5) Prerequisites(4) Motivation(4) Accessibility(4) Self-checking(3) Guidance(3) Context/application(2) Reasoning(1) Language(1)	Difficulty(27) Layout(21) Routine/drill(17) Problem solving(13) Accessibility(12) Motivation(10) Goal differentiation(10) Transparent structure(8) Presentation(7) Self-checking(7) Openness(7) Creativity(5) Size(5) Guidance(3) Support(2) Reasoning(1) Language(1) Context/application(1)	Difficulty(6) Routine/drill(5) Problem solving(4) Self-checking(3) Openness(3) Language(2) Accessibility(2) Layout(2) Goal Differentiation(2) Creativity(1) Transparent structure(1) Motivation(1) Guidance(1)

Table D: Assessment of Task 5.

Assessment of Task 5				
Number of mentions of task features				
All assessments	**Unsuitable**	**Rather unsuitable**	**Rather suitable**	**Suitable**
Routine/drill(42) Layout(21) Goal diff.(21) Difficulty(18) Motivation(9) Size(9) Transparent structure(6) Openness(5) Prerequisites(3) Language(3) Accessibility(3) Guidance(2) Support(2) Problem solving(1) Representation(1)	Routine/drill(13) Layout(6) Goal diff.(4) Transparent structure(3) Difficulty(3) Openness(3) Motivation(2) Problem solving(1) Size(1)	Routine/drill(19) Goal diff.(12) Difficulty(10) Layout(9) Size(4) Prerequisites(3) Transparent structure(2) Motivation(2) Guidance(2) Support(2) Representation(1) Openness(1)	Routine/drill(8) Layout(5) Motivation(4) Difficulty(4) Goal diff.(3) Language(2) Accessibility(2) Size(2) Openness(1)	Size(2) Routine/drill(2) Goal diff.(2) Transparent structure(1) Motivation(1) Language(1) Accessibility(1) Layout(1) Difficulty(1)