

Supporting student motivation in class: the motivational potential of tasks

Anna Heinle  · Anja Schiepe-Tiska · Frank Reinhold · Jörg-Henrik Heine · Doris Holzberger

Received: 6 May 2021 / Revised: 30 September 2021 / Accepted: 21 October 2021 / Published online: 28 April 2022
© The Author(s) 2022

Abstract Fostering student motivation is an important educational goal. However, motivation in the classrooms is rather heterogeneous, particularly in mathematics and physics. This study examines the potential of (textbook) tasks to promote student motivation. Based on *self-determination theory* (SDT) and *theory of interest*, a low-inference coding scheme was developed and validated by applying the framework of *item response theory* (IRT) to assess the motivational potential of tasks. Current ninth grade mathematics and physics tasks ($N=254$ task units) were analyzed using the categories *differentiated instruction*, *real-life context*, *autonomy support*, *competence support*, and *support for relatedness*. Additionally, differences between mathematics and physics tasks were examined. Results indicate the coding scheme's high interrater reliabilities and empirical validity. Furthermore, we found only a low occurrence of motivational features in mathematics and physics tasks, with few sub-

Anja Schiepe-Tiska is now at the University of Koblenz, Germany.

Anna Heinle (✉) · Prof. Dr. habil. Anja Schiepe-Tiska · Dr. phil. Jörg-Henrik Heine · Prof. Dr. phil. Doris Holzberger
Zentrum für internationale Vergleichsstudien (ZIB) e. V., TUM School of Education, Technische Universität München, Arcisstraße 21, 80333 München, Germany
E-Mail: anna.heinle@tum.de

Prof. Dr. habil. Anja Schiepe-Tiska
E-Mail: schiepe-tiska@uni-koblenz.de

Dr. phil. Jörg-Henrik Heine
E-Mail: joerg.heine@tum.de

Prof. Dr. phil. Doris Holzberger
E-Mail: doris.holzberger@tum.de

Prof. Dr. phil. Frank Reinhold
Institut für Mathematische Bildung Freiburg (IMBF), Pädagogische Hochschule Freiburg, Kunzenweg 21, 79117 Freiburg, Germany
E-Mail: frank.reinhold@ph-freiburg.de

ject-specific differences in favor of mathematics. The coding scheme can contribute to optimizing motivation-supportive instructional designs.

Keywords Mathematics · Motivational support · Physics · Task analysis

Förderung von Lernmotivation: das motivationsunterstützende Potential von Aufgaben

Zusammenfassung Die Förderung von Lernmotivation stellt ein wichtiges Bildungsziel dar. Gleichzeitig ist Lernmotivation im Unterricht heterogen ausgeprägt, besonders in den Fächern Mathematik und Physik. Dieser Beitrag untersucht das motivationsunterstützende Potential von (Schulbuch)-Aufgaben. Basierend auf theoretischen Arbeiten der *Selbstbestimmungstheorie* (SDT) und *Interesstheorie* wurde ein niedrig inferentes Kodierschema zur Erfassung des motivationalen Aufgabepotentials entwickelt und mithilfe der *Item Response Theory* (IRT) validiert. Dafür wurden aktuelle Mathematik- und Physikschulbuchaufgaben ($N=254$) der neunten Klasse anhand der Kategorien *innere Differenzierung*, *Lebensweltbezug*, *Autonomieunterstützung*, *Kompetenzunterstützung* und *soziale Einbindung* analysiert. Die Ergebnisse zeigen hohe Interraterreliabilitäten und eine empirische Validität des Kodierschemas. Darüber hinaus fanden wir eine geringe Ausprägung motivationaler Merkmale in Mathematik- und Physikaufgaben mit wenigen fachspezifischen Unterschieden zugunsten der Mathematik. Das entwickelte Kodierschema kann einen Beitrag zur Optimierung einer motivierenden Unterrichtsgestaltung leisten.

Schlüsselwörter Aufgabenanalyse · Motivationsförderung · Mathematik · Physik

1 Introduction

Student motivation is not only a crucial determinant for learning, but also a valuable educational goal itself (Schiepe-Tiska 2019). It represents an important condition for participating in society and for lifelong learning (OECD 2019). Within a classroom, student motivation can be very heterogeneous. It can be promoted through teacher characteristics (e.g., enthusiasm, Frenzel et al. 2009) and teacher behavior, (e.g., emotional support, Ruzek et al. 2016) and by providing specific learning opportunities such as motivation-enhancing tasks (Helmke 2009).

Particularly in STEM education, (textbook) tasks are at the center of learning processes (Knoll 2003). In mathematics, students spend about 80% of class time working on written tasks (Neubrand 2002) and about 82% of teachers use textbooks as a basis for instruction (Wendt et al. 2017). Previous approaches to analyzing the potential of tasks mainly focused on investigating cognitive features (e.g., Grünkorn et al. 2020; Jordan et al. 2008; Kühn 2010; Maier et al. 2010). Up to now, there have been only isolated attempts to identify tasks' motivational potential (Blömeke et al. 2006; Kassirra 2015).

The present study fills this gap and aims to 1) develop and validate a coding scheme for assessing the motivational potential of tasks, 2) examine tasks' motiva-

tional potential from current ninth grade German mathematics and physics textbooks, and 3) analyze the relationship between tasks' motivational potential and the level of task complexity.

2 Theoretical framework

2.1 Tasks as learning opportunities and their analyses

Tasks are written prompts for self-directed elaboration or exercise in student work phases (Kleinknecht 2019; Neubrand 2002). They reflect an important part of learning opportunities teachers provide. Tasks not only have the didactic function of performance testing, but additionally initiate and promote learning processes as well as review learning outcomes (Kleinknecht 2019). Above all, they help structure the learning environment (Meyer 2009).

Selecting and embedding tasks in classroom lessons are a challenge for teachers, as they need to assess the objective potential of a task (Neubrand 2002; Hammer 2015). A task's potential can be defined as "the opportunity for insightful learning inherent in the task that has not yet been realized" (Hammer 2015, p. 49). It can be assessed with the help of rational task analyses (Bromme et al. 1990).

In research, task analyses have a long tradition, especially in mathematics (e.g., Jordan et al. 2008; Neubrand 2002; OECD 2020) and science (Förtsch et al. 2018; Jatzwauk 2008; Kühn 2010), and in pedagogy (Maier et al. 2010). Previous research investigated the relation between objective task features and different learning outcomes (e.g., Förtsch et al. 2018; Jordan et al. 2008; OECD 2020). However, these studies focused primarily on the cognitive features of tasks. Nevertheless, mere cognitive demand is not enough for students' engagement with tasks. Also needed is motivational support (Stefanou et al. 2004), which can likewise be promoted through tasks (Jordan et al. 2008; Obersteiner et al. 2011).

2.2 The motivational potential of tasks

Motivation not only has an outstanding significance for learning and achievement (Krapp 2003), but is also an important basis for the willingness to engage in life-long learning (OECD 2019). Therefore, an important goal of education is student motivation (Aktionsrat Bildung 2015).

The general model of motivation proposes that current motivation arises from an interaction of person and situation factors (Heckhausen and Heckhausen 2018; Rheinberg and Vollmeyer 2019). In terms of promoting motivation, on the person side *self-determination theory* (SDT, Deci and Ryan 2000) is often used to explain how intrinsic motivation can be supported (Furtak and Kunter 2012), namely by fulfilling the basic needs for autonomy, competence, and relatedness. Experiences of autonomy arise when a person has the feeling of control and determination over their own behavior, goals, values, interests, and choices, even under certain constraints. The need for competence includes being effective, developing and evolving ones' own abilities, and experiencing progress in doing so. Relatedness refers to feeling

socially connected and cared for by others and belonging with significant others (e.g., Deci and Ryan 2000). The fulfillment of the three basic needs leads to self-determined learning.

The *theory of interest* expands on this and additionally considers interaction with situational factors by conceptualizing interest as a special relationship of a person to a specific object (e.g., certain activity, topic or content; see Krapp 2002). This interaction may lead to a current state of high situational interest (Krapp 2002). Teachers can stimulate situational interest by providing learning opportunities perceived as relevant for students' life by demonstrating their usefulness and necessity (Habig et al. 2018; Schraw 2001). In the context of learning, for current motivation it is additionally important to perceive a balance between ones' abilities and task demands (e.g., Csikszentmihalyi 2010; Eccles and Wigfield 2002). Following these approaches, we conclude that tasks as learning opportunities can represent a specific situation or object and thus have the potential to address basic needs and differentiate between levels of ability.

To date there has been little research on the motivational potential of tasks. Blömeke et al. (2006) presented a coding scheme that included some motivational features such as relevance of content, addressing individual needs, and social interaction. However, they applied it only to one mathematics example case task to illustrate how coding could work. In addition, Kassirra (2015) started to develop a coding scheme for *Work-Economy-Technology* tasks that included very detailed elaborations of motivational supportive features. However, his approach entailed methodological difficulties such as low interrater reliabilities among others, and he did not pursue it. Hence, an empirically validated low-inference coding scheme for assessing the motivational potential of mathematics and physics tasks has not yet been developed.

2.2.1 Developing a coding scheme

In line with SDT and *theory of interest*, Prenzel and Drechsel (1996) derived six empirically confirmed instructional conditions to promote self-determined learning that can be partially applied to the design of tasks: Teacher interest, content relevance, support for relatedness, autonomy support, competence support, and instructional quality with adapting difficulty to learners' prerequisites as an important aspect. In applying these conditions to the design of tasks, it is reasonable to focus on those conditions that learning opportunities themselves can provide, regardless of interactions with teachers in the classroom. Therefore, we propose five features to describe the motivational potential of tasks that are explained in more detail in the following: *Differentiated instruction*, *real-life context*, *autonomy support*, *competence support*, and *support for relatedness* (see Appendix 1). Further indicators and examples for these categories can be found in the full coding scheme, which is available on Open Science Framework (OSF): <https://osf.io/qxsjv/>.

Differentiated instruction One of the most prominent ideas for dealing with motivational heterogeneity in classrooms is differentiated instruction (e.g., Dumont 2019). Bönsch and Moegling (2012) distinguish five dimensions that can be ap-

plied to analyzing tasks and therefore have been incorporated in the coding scheme. (1) Tasks can differentiate in terms of goal structure and address different goals underlying an assignment. (2) Content structure refers to differentiating between levels of learners' prior knowledge. (3) Temporal structure means giving different possibilities depending on students' individual working speed. (4) In terms of action structure, tasks can enable students to apply different working techniques. (5) Tasks can support students to form social relationships in the classroom.

Real-life context Real-life context describes the relation between contents of the tasks and the real-life experience of students (Neubrand 2002). In accordance with Neubrand (2002) and Kleinknecht et al. (2011), three dimensions of real life context can be distinguished: (1) Tasks with constructed real life context hardly correspond to the experience of students, but combine the subject knowledge with a real-life context, albeit in an artificial way. (2) Authentic tasks contain context that seems to make sense in everyday life and is within the students' actual or real experiences and practices. (3) Within real life there are hardly any differences between the tasks problem and students' daily experience.

Autonomy support Tasks can nurture the need for autonomy by providing work assignments that are not prescriptive in detail but give choices that in turn lead to identifiable goals (Prenzel 1995). Furtak and Kunter (2012) distinguish three dimensions of autonomy support that can be applied to tasks' motivational potential: (1) Organizational autonomy support focuses on the surface structure and enables the choice of group members, assessment procedures, or one's own responsibility for deadlines. (2) Procedural autonomy support allows students to choose materials and equipment to process the task and provide the opportunity to demonstrate competencies in a freely chosen form. (3) Within cognitive autonomy supportive tasks, students may become initiators of their own learning, for example, by being empowered to find multiple solutions strategies to problems.

Competence support Tasks can meet the need for competence by (1) providing feedback through possibilities to check results or by asking for evaluation of others (Deci and Ryan 2002). Additionally and in accordance with Kassirra (2015), tasks have the potential to foster students' confidence to successfully develop competencies by (2) enabling students to actively demonstrate or present their competence and (3) to delve deeper into the subject of the task.

Support for relatedness Even though the satisfaction of this need also depends on classroom interaction, tasks can stimulate the feeling of relatedness by providing cooperative forms of work and learning (Deci and Ryan 2002). Following the work of Blömeke et al. (2006), they can initiate (1) working with partners or groups, or (2) discussing and reflecting on the results in class.

Although, Prenzel and Drechsel (1996) distinguish between instructional conditions as single constructs, in accordance with research on cognitive activation (Baumert et al. 2010; Herbert and Schweig 2021) and adaptive teaching (Dumont 2019) the proposed features may also reflect an overall construct of motivational

potential as an instructional strategy (i.e., a general factor). Hence, learning opportunities would support motivation not as individual features of tasks but as a whole (see Lazarides and Schiepe-Tiska 2021 for similar theoretical considerations).

2.2.2 *The relationship between task complexity and the motivational potential of mathematics and physics tasks*

Although we assume that motivational features within tasks are subject independent, tasks may also be determined by their context (Rakoczy 2008). The assessment of tasks' motivational potential is especially relevant in STEM education as students in Germany report below-average levels of enjoyment, interest, and motivation in mathematics and science (Schiepe-Tiska et al. 2016; Schiepe-Tiska and Schmidtner 2013). Within science, particularly physics is perceived as uninteresting (Prenzel et al. 2009b). Comparing the use of tasks in mathematics and physics instruction reveals differences. In mathematics, textbook tasks play a central role and are used by the majority of teachers, as the focus is primarily on practicing (Reiss and Hammer 2013). In physics, the focus is less on practicing but rather on experimenting and the teaching of scientific methods (Fraefel 2001). Nevertheless, experiments are usually embedded in tasks. Thus, they may play a less central but still important role. Physics teachers often use textbooks as supplements in their lessons (Wendt et al. 2017). These subject differences may result in differences in the motivational potential of tasks. Due to the strong focus on exercises in mathematics instruction (Rakoczy et al. 2019), one may assume feedback is present more in mathematics tasks (see also Hiebert et al. 2003). Also, it may seem plausible that in physics real-life context can be easier implemented in tasks due to its focus on experiments (Whitelegg and Parry 1999). However, the impact of a subject's context on tasks' motivational potential has not been examined yet.

Mathematics and physics tasks can comprise different levels of complexity describing the cognitive processes essential for successfully processing (Bloom 1972). The complexity of tasks may be related to their motivational potential. In line with German educational standards, we distinguish three levels of complexity: *reproduction*, *application*, and *transfer* (see methods section; Jordan et al. 2008; KMK 2003; Kühn 2010). All complexity levels can include different levels of difficulty, so there can be both easy transfer and difficult reproduction tasks (Jordan et al. 2008). There has been a massive amount of research on task complexity (e.g., Neubrand 2002; Robinson 2001; OECD 2020). However, its relationship with motivational potential has rarely been analyzed. Prenzel (1995) showed that for complex tasks, motivational support is particularly important. Tasks requiring higher cognitive performance achieve higher learning outcomes when they are embedded in self-determined forms of learning. For reproduction tasks, however, the motivational embedding was not related to higher learning outcomes.

3 Present study

We developed a coding scheme for the motivational potential of tasks and examined important aspects of its validity. In addition, we assessed the motivational potential of current ninth grade mathematics and physics textbook tasks and explored the relationship of tasks' motivational potential with their level of complexity.

To examine the validity of the coding scheme, we draw on the argumentation-based approach by Kane (2006). Arguments are formulated and evaluated for clarity, plausibility, and reasonableness of the inferences. Within this study, we focus on two related inference areas. *Generalization* includes assumption about the transferability of the instruments' results to comparable contexts. We assumed that the categories of motivational potential can be assessed in an intersubjective comprehensible way via the coding scheme (1.1). Moreover, we assumed that the sample used for validation represents the spectrum of current mathematics and physics tasks in a proper way (1.2). The second area, *scoring*, refers to assumptions about the scored data. We assumed that raters' understanding of the items is precise (1.3) and that the coding scheme's scoring rules are appropriate (1.4). Based on the theoretical framework of Prenzel and Drechsel (1996), we assumed that each of the five latent constructs differentiated *instruction*, *real-life context*, *autonomy support*, *competence support*, and *support for relatedness* can be represented by a unidimensional model (1.5.1). Moreover, based on theoretical considerations in the context of cognitive activation (Baumert et al. 2010; Herbert and Schweig 2021) and adaptive teaching (Dumont 2019), we tested whether the latent construct motivational potential is better be represented by a uni- or five-dimensional model (1.5.2).

To assess the motivational potential of current mathematics and physics textbook tasks, we examined how often the proposed motivational features were represented in tasks and investigated subject-specific differences. In line with current research (Rakoczy et al. 2019; Whitelegg and Parry 1999), we expect more feedback in mathematics tasks and more real-life context in physics tasks (2).

Last, we tested the relationship between tasks' motivational potential and their complexity level. Since higher level tasks lead to higher cognitive outcomes when they are integrated into motivational learning forms (Prenzel 1995), we expect higher motivational potential in more complex tasks (3).

4 Methods

4.1 Sampling procedure and sample description

Aiming at a proper representation throughout Germany, we sampled 200 tasks from current ninth grade mathematics ($n=100$) and physics textbooks ($n=100$) by using a top-down procedure (see Fig. 1). First, all 65 mathematics and 52 physics textbooks currently offered by the three largest German publishers were listed (issued between 2001 and 2020; $x=2014$). Besides publisher, the listed textbooks were clustered with regard to school type and federal state group. Although each German state has its own educational system, from the perspective of textbook design, they can be

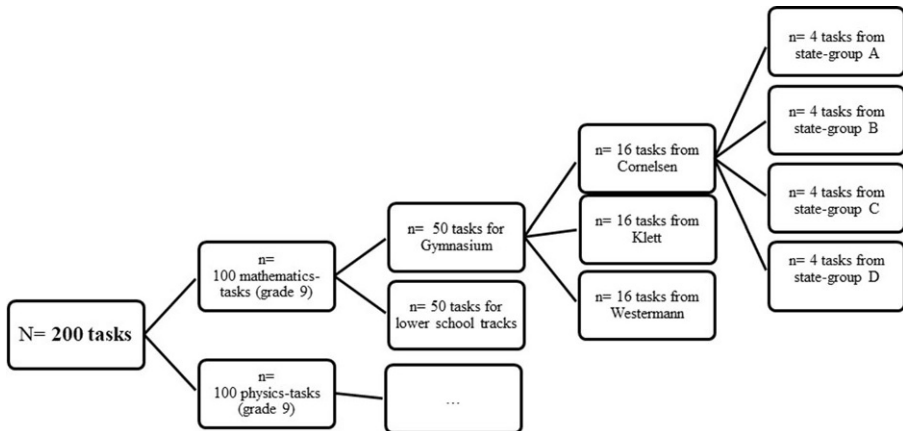


Fig. 1 Sampling structure. *Note.* state-group A: Bayern, Baden-Württemberg, Nordrhein-Westfalen; state-group B: Hessen, Rheinland-Pfalz, Saarland; state-group C: Schleswig-Holstein, Hamburg, Niedersachsen, Bremen; state group D: Sachsen, Sachsen-Anhalt, Mecklenburg-Vorpommern, Thüringen, Brandenburg, Berlin

divided into four target groups (see Fig. 1). Textbooks that were designed for all states were consequently listed in each cluster. Next, the average number of pages of the corresponding textbooks was determined for each federal state and school type cluster. We used 10% of the average number per cluster to select the textbook pages for the preliminary sample. For instance, if the average page number of a cluster is 250, every 25th page from each book listed was selected for the preliminary sample. Finally, four pages were randomly drawn from these clusters using SPSS statistical software (V. 27). To achieve as much variance as possible in task selection, an odd or even number was selected in alternation from the sampled textbook pages. If the selection criterion did not apply, the corresponding textbook was turned to the next page until the desired task number was available.

Textbook tasks often contained multiple subtasks. We treated each subtask having an independent or new work order as a single unit of analysis. In total, we sampled 254 units of analysis, 138 mathematics (Gymnasium: 70, lower school tracks: 68) and 116 physics units (Gymnasium: 63, lower school tracks: 53).

4.2 Coding scheme and rating procedure

The coding scheme was developed by content analysis method (e.g., Früh 2004). Categories were developed primarily deductively, i.e., driven by theory. One benefit of the deductive approach is that it reveals theoretical possibilities that remain unused in practice. Thus, possibilities for optimization become visible. In addition, one inductive category was added based on coding the text material, namely “choice of examples” as an indicator of autonomy support. By complementing the deductive with an inductive approach, categories also reflect indicators that are already established in real-life (Früh 2004).

We designed a low-inference coding scheme in which all features were coded using dichotomous response categories (applies or does not apply). All 254 units of analysis were double-coded by two coders, the first author and a trained research assistant. Inconsistencies were solved through regular discussions. To ensure a precise understanding of the rater, trainings took place at regular intervals including extensive explanations of subcategories and anchor examples, and repeated practice opportunities in the form of test coding were provided.

The coding scheme (see Appendix 1 and for a more detailed version OSF: <https://osf.io/qxsjv/>) consists of five categories: *differentiated instruction*, *real-life context*, *autonomy support*, *competence support*, and *support for relatedness* with two to five subcategories. Each subcategory is explained by a definition, multiple indicators, and anchor examples. At least one of the listed indicators had to be present in order to rate the category as applied.

Level of complexity was coded with three categories: reproduction, application, and transfer (Jordan et al. 2008; Kühn 2010, see Appendix 1 and for a more detailed version of OSF above). Again, all subcategories were coded using dichotomous response categories. For 38 task units, we observed more than one complexity level. These were recoded for the analyses and the highest complexity level was assigned to the task.

4.3 Statistical analyses

In order to assess whether the coding scheme can provide an intersubjectively comprehensible assessment of motivational potential, intercoder agreements were calculated. Therefore, the measure of exact percentage agreement was used. According to Stemler (2004) the minimum value is set at $\geq 75.00\%$. In order to correct for the possibility of agreement occurring by chance, Cohens' Kappa (κ) was computed (Wirtz and Caspar 2002). Following Wirtz and Caspar (2002), $0.75 < \kappa$ describes a very good, $0.6 < \kappa < 0.75$ describes a good, and $0.4 < \kappa < 0.6$ describes an acceptable concordance. Further, κ can take values from -1 to $+1$.

To test the dimensionality of the latent constructs, we applied the framework of *item response theory* (IRT). We used the package *pairwise* (Heine 2021) for R, which provides a non-iterative method for item parameter calibration and is suitable for small sample sizes. For each of the five latent constructs (i.e., the five categories of the coding scheme, see Appendix 1), we applied a unidimensional IRT scaling approach. In this case, the estimate of "person" measure represents the expression of a motivational category in a *task unit*. The global model fit was evaluated applying the Andersen Likelihood Ratio Test (Andersen 1973). In addition, the model fit statistics Q3 (Yen 1984; Christensen et al. 2017) were evaluated, which are based on residuals and have a recommended limit of $r_{Q3} < 0.2$ (see Christensen et al. 2017). Furthermore root-mean-square statistics (INFIT and OUTFIT) were calculated to assess item fit indices (Wright and Masters 1982). To test for the general construct *motivational potential*, we compared a unidimensional with a five-dimensional model estimated using *TAM* (Robitzsch et al. 2021).

To test differences between tasks' levels of complexity in their motivational potential, one-way ANOVAs were conducted. For the dependent variables, we used

the values of estimated person ability. Kolmogorov-Smirnov and Shapiro-Wilk indicated that the data were not normally distributed ($\alpha=0.05$). Hence, we used the more robust Welch-ANOVA (Welch 1947). The analyses scripts are available via OSF (for link, see above).

5 Results

5.1 Validity: Generalization

Agreement percentages for all items ranged from 92.9 to 100% and thus exceeded the minimum level of agreement (Stemler 2004, see Appendix 2). Cohen's kappa ranged from 0.76 to 1, with the value for *demonstration and presentation of competence* being the lowest. For the items *process*, *action*, and *social structure* as well as *organizational autonomy support* no kappa values could be calculated as these items were not coded in any task unit by either rater. For *cognitive autonomy support* the agreement was 99.2%. However, a negative kappa score was calculated due to the rare occurrence of the category, which resulted in uneven marginal sum distributions (Spitznagel and Helzer 1985). In sum, overall interrater reliability was very good. (H 1.1). The multi-level sampling design indicated a proper representation of current textbook tasks (H 1.2).

5.2 Validity: Scoring

Successfully completed rater trainings led to precise item understanding (H 1.3), which was reflected in a precise mapping of the items in the tasks due to appropriate coding rules (H 1.4). When testing the unidimensionality of the five latent constructs, we observed that six items *process*, *action*, and *social structure*, *organizational* and *cognitive autonomy support*, and *discussion and reflection in class* were not rated at all or at most once (see Appendix 3). Within the IRT scaling approach, the difficulty of these items represents a constant and thus, they had to be removed from the models. Hence, only the constructs *differentiated instruction*, *real-life context*, *autonomy support*, and *competence support* and the general factor *motivational support* could be scaled. For global model fit, the Andersen test showed no significant deviation from the model fit for all four scales, respectively, or for the general factor scale *motivational potential*, suggesting an overall fit of the scaling model for the five scales. Contrary to the findings from the Andersen test, coefficients for the Yen's Q3 statistics for *real-life context* ($r_{Q3}=0.299$), *competence support* ($r_{Q3}=0.266$) and *motivational potential* ($r_{Q3}=0.319$) fell above the recommended limit (see Christensen et al. 2017; see Appendix 4). The coefficients for the item root-mean-square fit indices (MSQ) consistently showed values for OUTFIT and INFIT smaller than the expected value of $MSQ_{exp}=1$ (see Appendix 5). This indicates a model overfit, which means the model fits the Rasch model better than expected (Wright and Masters 1982) (1.5.1). Comparing a model representing four latent constructs with a unidimensional model, the Bayes Information Criterion (BIC) index of a unidimensional model was smaller as the BIC of the four-dimen-

sional model (see Schwarz 1978; see Appendix 6). Thus, *motivational potential* can represent a unidimensional factor (1.5.2).

5.3 Motivational potential of mathematics and physics tasks

Appendix 3 shows that, overall, the motivational potential in mathematics and physics tasks was rather low. Some of the categories were not coded at all resulting in a skewed distribution of the items. Hence, we reject further statistical analyses and report frequencies only.

For *differential instruction*, only three subcategories were coded for mathematics and physics tasks, namely *goal*, *content*, and *action structure*, whereby *content structure* was coded more frequently in mathematics tasks (in approx. 12.3% of all mathematics tasks and in 4.3% of all physics tasks). *Action structure* was observed only in one mathematics task.

Real-life context occurred equally often in both subjects. A *constructed life context* was most common (in approx. 28% of all tasks), while *real-life tasks* were least common (in approx. 5.5% of all tasks). However, about half of the tasks did not display any real life-context.

Autonomy support was hardly present in either subsample. *Organizational autonomy support* was not coded at all and *cognitive autonomy support* was coded only once in a physics task. *Procedural autonomy support* was present in both subjects in about 1.5% of all tasks. The most frequent category in both subjects was *choice of examples* in about 6% of all tasks.

Competence support was also hardly coded. Noticeably, a concrete reference for *feedback* was given only in mathematics tasks (approx. 7.2%). References for *demonstrating or presenting results* or for *further deliberation on task content* occurred in both subjects about equally seldom (approx. 3.5%).

A similar picture emerged for *support for relatedness*. In the mathematics tasks, specific references for *partner and group work* appeared more frequently (approx. 7.2%). Overall, the explicit *opportunity to reflect in the classroom* was coded only in one physics task.

5.4 Relation between motivational potential and level of complexity

Because we found hardly any differences between physics and mathematics tasks, in the following we do not distinguish between subjects. Moreover, as the model comparison of dimensionality revealed a general factor, we focus on the overall *motivational potential*.

Motivational potential differed statistically significantly for the different levels of complexity in tasks with a small effect, Welch's $F(2, 90.71) = 10.18$, $p < 0.001$, $\eta^2 = 0.057$. Games-Howell post-hoc analysis revealed a significant difference ($p < 0.001$) between reproduction ($M = -2.85$) and application tasks ($M = -2.51$) in favor of application tasks (-0.34 , 95%-CI $[-0.61, -0.07]$) and a difference between reproduction and transfer tasks ($M = -2.22$) in favor of transfer tasks (-0.63 , %-CI $[-0.98, -0.29]$).

6 Discussion

The present study developed and validated a coding scheme to assess tasks' motivational potential and applied it to current ninth grade German mathematics and physics textbook tasks. In addition, the relation between tasks' motivational potential and their complexity level was analyzed. Key results were: 1) Overall, the developed coding scheme could be evaluated as valid. 2) For current mathematics and physics textbook tasks, only low motivational potential was found, with *real-life context* being present most frequently. 3) There were only minor subject differences. 4) The overall *motivational potential* was higher in application and transfer tasks than in reproduction tasks.

6.1 Coding and designing the motivational potential of tasks

The validity testing of the coding scheme indicated a precise rater understanding, as all items could be clearly assigned to all tasks due to appropriate scoring rules. This was also reflected in high interrater reliabilities, showing that the assessment with the coding scheme could be generalized across both raters. Within the statistical validation of the theoretically derived motivational features, *differentiated instruction* and *autonomy support* could be scaled with acceptable model fits. In line with research on cognitive activation (Baumert et al. 2010; Herbert and Schweig 2021) and adaptive teaching (Dumont 2019), we could confirm that *motivational potential* in tasks can be considered as an overall strategy and thus is not limited to individual features.

The coding scheme could therefore be used in teacher education and training to raise awareness on which tasks features might be motivational. Furthermore, it could support teachers in planning and preparation lessons to identify the motivational potential of a task in order to optimize if necessary. Finally, the coding scheme could assist textbook publishers in designing motivational tasks.

6.2 Motivational potential of mathematics and physics tasks and the relation with level of complexity

The overall low frequency of motivational features in mathematics and physics textbook tasks was somewhat surprising, as Germany put some effort in implementing a new or more developed task culture in the STEM field by introducing the SINUS program (Prenzel et al. 2009a) or the *Context-projects* (e.g., Mikelskis-Seifert and Duit 2007) that also aimed at incorporating motivational features such as differentiated instruction and real-life context (Horn 1999). In line with these efforts, in our study, *real-life context* was observed most frequently and it was similarly distributed in both subjects. However, we would have expected a higher frequency in physics tasks as, in theory, physics can offer more room for implementing real-life experiences. Yet, this is consistent with other empirical findings showing only limited implementation of real-life context in physics classrooms (Haag and Götz 2012). Nevertheless, a task's objective real-life context does not guarantee its connection to students' actual and strongly individual experiences.

Overall, subject context played only a minor role for the motivational potential of tasks.

Notable differences emerged only for the subcategory *feedback* of the construct *competence support* in favor of mathematics. This reflects that practicing and thus monitoring play a greater role in mathematics than in science (Fraefel 2001). A second notable result is that, although *relatedness* has shown to be a key function for longer-term interest in physics (Hazari et al. 2010) and that it can be promoted especially when conducting experiments, only mathematics tasks included specific references to partner and group work. However, maybe teachers do encourage partner and group work in classroom interaction without explicit references in textbook tasks.

With regard to task complexity, as expected, complex tasks contained higher motivational potential and thus offered more options for processing the assignment. Still, the difference was rather small. Although motivational features within reproduction tasks do not seem to be in relation with high learning outcomes (Prenzel 1995), we believe they are nevertheless important for some individuals, especially for their motivational outcomes, which themselves are important learning goals. There are already textbook concepts that focus on the integration of motivational task features for all complexity levels, mainly for vocational schools (see Riecke-Baulecke and Broux 2016) that could be used as an orientation for future textbook developments for STEM education of other school types.

6.3 Limitations and future research

The present paper extends previous task analysis research by focusing on the motivational potential of tasks. Moreover, a low-inference coding scheme was developed and empirically validated using the IRT scaling approach, with the person estimator representing the expression of a motivational category in a task unit. For this, 254 mathematics and physics textbook tasks units were sampled using a top-down procedure and thus reflect a proper representation throughout Germany. Nevertheless, some limitations should be considered.

First, this study focused on tasks analysis and therefore did not test all of Kane's recommended validity arguments (Kane 2006). Subsequent research should consider additional areas of inference, such as extrapolation, to compare the coding scores with other assessments of *motivational potential* (see Herbert and Schweig 2021).

Second, examining current mathematics and physics textbook tasks revealed the problem of skewed item value distributions for some constructs. Low item frequencies caused misbehavior and made further statistical analyses difficult. However, we do not attribute the low item frequencies to the scoring design of the coding scheme, as for all items anchor examples from mathematics and physics textbooks could be found. Rather, we believe that current mathematics and physics tasks, which were represented properly within this study, have explicit shortcomings in motivational features. Future studies with an expanded sample size are needed to examine whether this would increase items' expression. Tasks from other domains, such as language, could be used to expand research on the scaling of coding schemes and further test the context independency of tasks' motivational potential.

Third, the current study focused exclusively on tasks as learning opportunities and their motivational potential. However, theories of instruction such as the Generic Dimensions of Instruction (e.g., Holzberger et al. 2019; Praetorius et al. 2018) show that motivational-affective outcomes can also be promoted through effective socio-emotional support, which depends on teacher-student interactions. Thus, we believe that within instruction the motivational potential of tasks will spread its effect by supplementing teacher characteristics and behavior (e.g., Frenzel et al. 2009; Ruzek et al. 2016). This interaction could be explored further in future studies.

To examine how the motivational potential of a task can actually enhance motivation, we suggest that future studies supplement task analyses with (video) observations. By doing so, one could analyze how teachers embed tasks in their instruction (as proposed by Blömeke et al. 2006 and Rakoczy 2008) and the perception and use of tasks by students (Helmke 2009). Along with this, teacher competencies such as diagnostic competence should be considered (Südkamp and Praetorius 2017). It would be also important, to clarify how tasks' motivational potential interacts with students' motivational orientations and needs (Vansteenkiste et al. 2009).

7 Conclusion

By using a primarily deductive approach, we developed and validated a coding scheme to identify theoretical opportunities for the motivational potential of tasks. Our results showed that in current German ninth grade mathematics and physics textbooks the possibilities for promoting student motivation remain unexploited and leave room for improvement. In addition, our low-inference coding scheme can be used not only in future research but also in teacher training and teaching practice to assess the motivational potential of tasks and to design motivational instruction.

Supplementary Information The online version of this article (<https://doi.org/10.1007/s11618-022-01090-3>) contains supplementary material, which is available to authorized users.

Funding This research was funded by the Federal Ministry of Education and Research (BMBF) and the Standing Conference of the Ministers of Education and Cultural Affairs of the Federal Republic of Germany (KMK) (ZIB, 2022).

Funding Open Access funding enabled and organized by Projekt DEAL.

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

Conflict of interest A. Heinle, A. Schiepe-Tiska, F. Reinhold, J.-H. Heine and D. Holzberger declare that they have no competing interests.

References

- Aktionsrat Bildung (2015). *Bildung. Mehr als Fachlichkeit. Gutachten*. Münster: Waxmann.
- Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26(1), 31–44. <https://doi.org/10.1111/j.2044-8317.1973.tb00504.x>.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>.
- 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.
- Bloom, B. S. (1972). *Taxonomie von Lernzielen im kognitiven Bereich (Beltz-Studienbuch)*. Weinheim: Beltz.
- Bönsch, M., & Moegling, K. (Eds.). (2012). *Binnendifferenzierung. 2. Unterrichtsbeispiele für den binnendifferenzierten Unterricht* (Theorie und Praxis der Schulpädagogik, Vol. 18). Immenhausen: Prolog.
- Bromme, R., Seeger, F., & Steinbring, H. (1990). *Aufgaben als Anforderungen an Lehrer und Schüler* (IDM-Untersuchungen zum Mathematikunterricht, Vol. 14). Köln: Aulis-Verlag Deubner.
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for yen's Q3: identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement*, 41(3), 178–194. <https://doi.org/10.1177/0146621616677520>.
- Csikszentmihalyi, M. (2010). *Das Flow-Erlebnis. Jenseits von Angst und Langeweile: im Tun aufgehen* (11th edn., Konzepte der Humanwissenschaften). Stuttgart: Klett-Cotta.
- Deci, E., & Ryan, R. (2000). The "what" and "why" of goal pursuits: Human needs and the Self-Determination of behavior. *Psychological Inquiry*, 11, 227–268. https://doi.org/10.1207/S15327965PLI1104_01.
- Deci, E. L., & Ryan, R. M. (Eds.). (2002). *Handbook of Self-Determination research*. Rochester: University of Rochester Press.
- Dumont, H. (2019). Neuer Schlauch für alten Wein? Eine konzeptuelle Betrachtung von individueller Förderung im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 22(2), 249–277. <https://doi.org/10.1007/s11618-018-0840-0>.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>.
- Förtsch, S., Förtsch, C., von Kotzebue, L., & Neuhaus, B. J. (2018). Effects of teachers' professional knowledge and their use of three-dimensional physical models in biology lessons on students' achievement. *Education Sciences*. <https://doi.org/10.3390/educsci8030118>.
- Fraefel, U. (2001). *Merkmale und Scripts des Naturwissenschaftsunterrichts der Sekundarstufe I in der Deutschschweiz: Analyse auf der Grundlage videographierter Lektionen*. Zürich: Pädagogisches Institut der Universität.
- Frenzel, A. C., Goetz, T., Lüdtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology*, 101(3), 705–716. <https://doi.org/10.1037/a0014695>.
- Früh, W. (2004). *Inhaltsanalyse: Theorie und Praxis*. Konstanz: UVK Verlagsgesellschaft mbH.
- Furtak, E., & Kunter, M. (2012). Effects of autonomy-supportive teaching on student learning and motivation. *Journal of Experimental Education*, 80(3), 284–316. <https://doi.org/10.1080/00220973.2011.573019>.
- Grünkorn, J., Klieme, E., Praetorius, A.-K., & Schreyer, P. (Eds.). (2020). *Mathematikunterricht im internationalen Vergleich: Ergebnisse aus der TALIS-Videostudie Deutschland*. Frankfurt am Main: DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation.
- Haag, L., & Götz, T. (2012). Mathe ist schwierig und Deutsch aktuell: Vergleichende Studie zur Charakterisierung von Schulfächern aus Schülersicht. *Psychologie in Erziehung und Unterricht*, 59(1), 32–46. <https://doi.org/10.2378/peu2012.art03d>.
- Habig, S., van Vorst, H., & Sumfleth, E. (2018). Merkmale kontextualisierter Lernaufgaben und ihre Wirkung auf das situationale Interesse und die Lernleistung von Schülerinnen und Schülern. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1), 99–114. <https://doi.org/10.1007/s40573-018-0077-8>.

- Hammer, S. (2015). *Professionelle Kompetenz von Mathematiklehrkräften im Umgang mit Aufgaben in der Unterrichtsplanung: Theoretische Grundlegung und empirische Untersuchung* (Dissertation). München: Ludwig-Maximilians-Universität.
- Hazari, Z., Sonnert, G., Sadler, P.M., & Shanahan, M.-C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: a gender study. *Journal of Research in Science Teaching*, 47(8), 978–1003. <https://doi.org/10.1002/tea.20363>.
- Heckhausen, J., & Heckhausen, H. (Eds.). (2018). *Motivation and action* (3rd edn.). Cham: Springer. <https://doi.org/10.1007/978-3-319-65094-4>.
- Heine, J.-H. (2021). pairwise: Rasch model parameters by Pairwise algorithm. Computer software. <https://cran.r-project.org/package=pairwise>. Accessed 26 Apr 2021.
- Helmke, A. (Ed.). (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze-Velber: Kallmeyer. Neubearbeitung
- Herbert, B., & Schweig, J. (2021). Erfassen des Potenzials zur kognitiven Aktivierung über Unterrichtsmaterialien im Mathematikunterricht. *Zeitschrift für Erziehungswissenschaft*. <https://doi.org/10.1007/s11618-021-01020-9>.
- Hiebert, J., Gallimore, R., Garnier, K., Givvin, K., Hollingsworth, H., Jacobs, J., Angel, M.-Y.C., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (Eds.). (2003). *Teaching mathematics in seven countries: results from the TIMSS 1999 video study*. Washington DC: U.S. Department of Education, National Center for Education Statistics.
- Holzberger, D., Praetorius, A.-K., Seidel, T., & Kunter, M. (2019). Identifying effective teachers: the relation between teaching profiles and students' development in achievement and enjoyment. *European Journal of Psychology of Education*, 34(4), 801–823. <https://doi.org/10.1007/s10212-018-00410-8>.
- Horn, M.E. (1999). *Klassifizierung von Physikaufgaben: Materialsichtung und Materialentwicklung für das Pädagogische Landesinstitut Brandenburg im Rahmen des BLK-Programms SINUS*.
- Jatzwauk, P. (2008). Der Einfluss des Aufgabeneinsatzes im Biologieunterricht auf die Lernleistung der Schüler: Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 263–283.
- Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., Kunter, M., & Baumert, J. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht. *Journal für Mathematik-Didaktik*, 29(2), 83–107.
- Kane, M. (2006). Content-related validity evidence in test development. In S.M. Downing & T.M. Haladyna (Eds.), *Handbook of test development* (pp. 131–153). Mahwah: Lawrence Erlbaum.
- Kassirra, R. (2015). Ein methodisches Verfahren zur Analyse motivationsunterstützender Merkmale in Arbeitsaufträgen. In Gesellschaft für Arbeitswissenschaft e.V. (GfA) (Ed.), *VerANTWORTung für die Arbeit der Zukunft. Band zum 61. Kongress der Gesellschaft für Arbeitswissenschaft* (25. bis 27. Februar 2015, Beitrag B.1.10). Dortmund: Gesellschaft für Arbeitswissenschaft e.V.
- Kleinknecht, M. (2019). Aufgaben und Aufgabenkultur. *Zeitschrift für Grundschulforschung*, 12(7), 1–14. <https://doi.org/10.1007/s42278-018-00035-2>.
- Kleinknecht, M., Bohl, T., Maier, U., & Metz, K. (2011). Aufgaben und Unterrichtsplanung. In K.-H. Arnold, T. Bohl & K. Zierer (Eds.), *Jahrbuch für Allgemeine Didaktik 2011* (pp. 59–75). Hohengehren: Schneider-Verlag.
- KMK (2003). Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 04.12.2003. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2003/2003_12_04-Bildungsstandards-Mathe-Mittleren-SA.pdf. Accessed 26 Apr 2021.
- Knoll, S. (2003). *Verwendung von Aufgaben in Einführungsphasen des Mathematikunterrichts*. Marburg: Tectum.
- Krapp, A. (2002). An educational-psychological theory of interest and its relation to SDT. In E.L. Deci & R.M. Ryan (Eds.), *Handbook of self-determination research* (pp. 405–427). Rochester: University of Rochester Press.
- Krapp, A. (2003). Die Bedeutung der Lernmotivation für die Optimierung des schulischen Bildungssystems. *Politische Studien*, 54(3), 91–105.
- Kühn, S.M. (Ed.). (2010). *Steuerung und Innovation durch Abschlussprüfungen*. Wiesbaden: VS. <https://doi.org/10.1007/978-3-531-92418-2>.
- Lazarides, R., & Schiepe-Tiska, A. (2021). Heterogenität motivationaler Merkmale im Unterrichtskontext. *Zeitschrift für Erziehungswissenschaft*. <https://doi.org/10.1007/s11618-022-01082-3>.
- 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, 84–96.
- Meyer, H. (2009). *Was ist guter Unterricht?* Berlin: Cornelsen.

- Mikelskis-Seifert, S., & Duit, R. (2007). Physik im Kontext – Innovative Unterrichtsansätze für den Schullalltag. *MNU*, 60(5), 265–274.
- Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen: Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMMS-Video-Studie* (Dissertation). Hildesheim: Franzbecker.
- Obersteiner, A., Reiss, K., & Martel, A. (2011). Offene Aufgaben in Schulbüchern und ihr Einsatz im Mathematikunterricht. In E. Matthes & S. Schütze (Eds.), *Aufgaben im Schulbuch* (Beiträge zur historischen und systematischen Schulbuchforschung, pp. 303–313). Bad Heilbrunn: Klinkhardt.
- OECD (2019). *OECD future of education and skills 2030. OECD learning compass 2030*. Paris: OECD Publishing.
- OECD (2020). *Global teaching InSights. A video study of teaching*. Paris: OECD.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*, 50(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>.
- Prenzel, M. (1995). Zum Lernen bewegen: Unterstützung von Lernmotivation durch Lehre. *Blick in die Wissenschaft Forschungsmagazin der Universität Regensburg*, 4(7), 58–71.
- Prenzel, M., & Drechsel, B. (1996). Ein Jahr kaufmännische Erstausbildung: Veränderungen in Lernmotivation und Interesse. *Unterrichtswissenschaft*, 24(3), 217–234.
- Prenzel, M., Friedrich, A., & Stadler, M. (Eds.). (2009a). *Von SINUS lernen – wie Unterrichtsentwicklung gelingt* (1st edn., Sinus-Transfer). Seelze-Velber: Klett.
- Prenzel, M., Reiss, K., & Hasselhorn, M. (2009b). Förderung der Kompetenzen von Kindern und Jugendlichen. In J. Milberg (Ed.), *Förderung des Nachwuchses In Technik und Naturwissenschaft: Beiträge zu den Zentralen Handlungsfeldern* (acatech DISKUTIERT, pp. 15–46). Berlin: Springer. https://doi.org/10.1007/978-3-642-01123-8_1.
- Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht. Unterricht aus der Perspektive von Lernenden und Beobachtern* (Pädagogische Psychologie und Entwicklungspsychologie, Vol. 65). Münster: Waxmann.
- Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction*, 60, 154–165. <https://doi.org/10.1016/j.learninstruc.2018.01.004>.
- Reiss, K., & Hammer, C. (2013). *Grundlagen der Mathematikdidaktik*. Basel: Springer Basel. <https://doi.org/10.1007/978-3-0346-0647-9>.
- Rheinberg, F., & Vollmeyer, R. (2019). *Motivation* (9th edn., Kohlhammer-Urban-Taschenbücher, Vol. 555). Stuttgart: Kohlhammer.
- Riecke-Baulecke, T., & Broux, A. (Eds.). (2016). *Basistrainer Mathematik, Deutsch und Englisch zur Vorbereitung auf Ausbildung und Beruf* (1st edn.). Köln: Bildungsverlag EINS.
- Robinson, P. (2001). Task complexity, task difficulty, and task production: exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. <https://doi.org/10.1093/applin/22.1.27>.
- Robitzsch, A., Kiefer, T., & Wu, M. (2021). TAM: test analysis modules. <https://cran.r-project.org/web/packages/TAM/index.html>. Accessed 3 May 2021.
- Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2016). How teacher emotional support motivates students: the mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction*, 42, 95–103. <https://doi.org/10.1016/j.learninstruc.2016.01.004>.
- Schiepe-Tiska, A. (2019). *Mehrdimensionale Bildungsziele in internationalen Large-Scale Assessments: Konzeptualisierung, Entwicklung und Entstehungsbedingungen am Beispiel der MINT-Fächer* (Habilitation). München: Technische Universität München.
- Schiepe-Tiska, A., & Schmidtnr, S. (2013). Mathematikbezogene emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen von Jugendlichen in PISA 2012. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Eds.), *Pisa 2012: Fortschritte und Herausforderungen in Deutschland* (pp. 99–122). Münster: Waxmann.
- Schiepe-Tiska, A., Schmidtnr, S., Müller, K., Heine, J.-H., Neumann, K., & Lütke, O. (2016). Naturwissenschaftlicher Unterricht in PISA 2015 im internationalen Vergleich. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Eds.), *PISA 2015: Eine Studie zwischen Kontinuität und Innovation* (pp. 133–175). Münster: Waxmann.
- Schraw, G. (2001). Promoting general metacognitive awareness. In H. J. Hartman (Ed.), *Metacognition in learning and instruction: theory, research and practice* (pp. 3–16). Dordrecht: Springer. https://doi.org/10.1007/978-94-017-2243-8_1.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.

- Spitznagel, E.L., & Helzer, J.E. (1985). A proposed solution to the base rate problem in the kappa statistic. *Archives of General Psychiatry*, 42(7), 725–728. <https://doi.org/10.1001/archpsyc.1985.01790300093012>.
- Stefanou, C.R., Perencevich, K.C., DiCintio, M., & Turner, J.C. (2004). Supporting autonomy in the classroom: ways teachers encourage student decision making and ownership. *Educational Psychologist*, 39(2), 97–110. https://doi.org/10.1207/s15326985ep3902_2.
- Stemler, S.E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation*. <https://doi.org/10.7275/96jpxz07>.
- Südkamp, A., & Praetorius, A.-K. (Eds.). (2017). *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Vol. 94). Münster: Waxmann.
- Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: the quality of motivation matters. *Journal of Educational Psychology*, 101(3), 671–688. <https://doi.org/10.1037/a0015083>.
- Welch, B.L. (1947). The generalization of ‘Student’s’ problem when several different population variances are involved. *Biometrika*, 34(1/2), 28–35. <https://doi.org/10.2307/2332510>.
- Wendt, H., Bos, W., Goy, M., & Jusufi, D. (Eds.). (2017). *TIMSS 2015: Skalenhandbuch zur Dokumentation der Erhebungsinstrumente und Arbeit mit den Datensätzen*. Münster: Waxmann.
- Whitelegg, E., & Parry, M. (1999). Real-life contexts for learning physics: meanings, issues and practice. *Physics Education*, 34(2), 68–72. <https://doi.org/10.1088/0031-9120/34/2/014>.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA Press.
- Yen, W.M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125–145. <https://doi.org/10.1177/014662168400800201>.