Journal für Mathematik-Didaktik (2025) 46:12
https://doi.org/10.1007/s13138-025-00263-x

ORIGINALARBEIT/ORIGINAL ARTICLE

Highlight, Write, Elaborate: Note-Taking Strategies to Master Reality-Based Mathematical Tasks

Lisa-Marie Wienecke · Dominik Leiss · Timo Ehmke

Received: 7 January 2025 / Accepted: 28 July 2025 / Published online: 1 September 2025

Abstract This study investigates the effect of note-taking strategies on solving reality-based mathematical tasks, such as highlighting relevant information, writing relevant data, and elaborating on information during the solution process. While prior research has highlighted the general benefits of note-taking for learning, few studies have examined how specific note-taking strategies operate within the context of complex, reality-based mathematics tasks. This study extends existing work by systematically analyzing how different note-taking subprocesses interact with student characteristics and influence solving success. Data of 1064 task solutions from students in grades 7–10 ($M_{age} = 14.85$, $SD_{age} = 1.26$) were collected from three task contexts, each at three levels of language, and coded for subprocesses, namely, relevant and irrelevant notes, calculation paths, and elaboration strategies. The methodology included descriptive statistics, correlation analyses corrected for multiple comparisons and generalized linear mixed models to explore the relationship between note-taking strategies, personal factors (e.g., mathematical and language proficiency), and task performance. Effective note-taking strategies, such as writing and elaborating relevant information, significantly contributed to correct task solutions. Conversely, taking irrelevant notes was negatively associated with performance. The generalized linear mixed models indicated that personal characteristics, including general mathematics and language proficiency, predicted task success, text understanding, and various note-taking types. Effective note-taking enhanced students' capacity to tackle complex problems. These findings address previous calls to examine in detail the subprocesses of solution strategies and underscore the importance of teaching different note-taking strategies to enhance the solving process in reality-based tasks.

✉ Lisa-Marie Wienecke · Dominik Leiss
Institute of Mathematics and its Didactics, Leuphana University Lueneburg, Lueneburg, Germany
E-Mail: lisa-marie.wienecke@leuphana.de

Timo Ehmke
Institute for Educational Sience, Leuphana University Lueneburg, Lueneburg, Germany

Keywords Mathematical modelling · Elaboration strategies · Word problems · Note-taking · Organization strategies · Language competence

Hervorheben, Herausschreiben, Elaborieren: Notizenstrategien um den Herausforderungen des Lösens realitätsbezogener Aufgaben zu begegnen

Zusammenfassung Diese Studie untersucht die Zusammenhänge zwischen den Notizenstrategien im Lösungsprozess, wie dem Hervorheben relevanter Informationen, dem Aufschreiben relevanter Daten und dem Elaborierend von Informationen, und dem Lösen realitätsbezogener Aufgaben. Während frühere Untersuchungen das Anfertigen von Notizen für das Lernen hervorgehoben haben, gibt es nur wenige Studien, die untersucht haben, wie bestimmte Notizen im Kontext des Lösungsprozesses realitätsbezogener Mathematikaufgaben wirken. Die vorliegende Studie erweitert den bestehenden Forschungsstand, indem sie systematisch analysiert, wie verschiedene Teilprozesse, wie beispielsweise das Hervorheben zur Lösung benötigter Informationen und das Elaborieren jener, mit den Merkmalen der Lernenden interagieren und der Aufgabenleistung beeinflussen. Die Daten von 1064 Aufgabenlösungen von Lernenden der Klassen 7 bis 10 ($_{age}$ = 14,85, SD_{age} = 1,26) wurden anhand von drei Aufgabenkontexten mit jeweils drei Sprachniveaus erhoben. Die Analysen umfassen deskriptive Statistiken, korrigierte Korrelationsanalysen und lineare gemischte Modelle (GLMM), um den Zusammenhang zwischen Notizstrategien, Personenmerkmalen (z. B. mathematische und sprachliche Kompetenzen) und der Aufgabenleistung zu untersuchen. Strategien, wie das Aufschreiben und Elaborieren relevanter Informationen, trugen signifikant zur korrekten Lösung der Aufgaben bei. Umgekehrt ist das Notieren irrelevanter Informationen negativ mit der Leistung verbunden. Die GLMM zeigten, dass persönliche Merkmale, darunter allgemeine Mathematik- und Sprachkenntnisse, das Textverständnis und verschiedene Arten der Notizen den Erfolg beim Lösen der Aufgabe vorhersagen. Diese Ergebnisse entsprechen früheren Forderungen, die Teilprozesse von Lösungsstrategien im Detail zu untersuchen, und unterstreichen die Bedeutung des Unterrichtens verschiedener Notizen-Strategien, um den Lösungsprozess bei realitätsnahen Aufgaben zu verbessern.

Schlüsselwörter Mathematisches Modellieren · Elaborationsstrategien · Realitätsbezogene Aufgaben · Notizen · Organisationsstrategien · Sprachkompetenz

1 Introduction

Preparing students for real-life requirements is a core objective of schooling (Gravemeijer et al. 2017; National Council of Teachers of Mathematics 2000; Organisation for Economic Co-operation and Development 2013). To address 21st-century challenges, mathematics education must be contextualized for interdisciplinary, real-world demands (Goos et al. 2023). Scientific—especially mathematical—literacy should form the foundation of contemporary research (Härtig et al. 2015). However, traditional instruction often relies on decontextualized tasks with limited practical relevance (Verschaffel et al. 2000). This calls for a reform in task design (Stein et al. 1996), and researchers have responded by exploring context-rich, reality-based tasks (Boaler 2001; Van Den Heuvel-Panhuizen 2003). These tasks aim to foster authenticity (Kaiser 2017; Schlüter and Besser 2024) and better prepare students for applying mathematics in everyday life (Gravemeijer et al. 2017; Niss et al. 2007).

Moreover, real-world tasks often show lower solution rates than traditional ones (Daroczy et al. 2015; De Bock et al. 2003). More than two decades after the first “PISA shock” (Boaler 2001), the 2022 PISA results once again exposed the fragility of students’ basic mathematical competencies. Despite earlier improvements in Germany between 2000 and 2012, current results reveal a sharp decline, with mathematical performance reaching a new low in 2022 (Lewalter et al. 2023). Additionally, nearly a quarter of German students demonstrated only limited reading proficiency—posing serious challenges for understanding and solving text-based mathematical problems (Lewalter et al. 2023). These findings underscore the need to support students in navigating the linguistic and cognitive complexity of real-world mathematical tasks. El Masri et al. (2017) note that existing models explain only 20–23% of performance variance in such tasks, suggesting further research.

A promising but underexplored angle is students’ strategic behavior during solving reality-based tasks. While various strategies have been discussed (e.g., modelling heuristics; Leiss and Tropper 2014), note-taking offers a particularly observable and practical behavior. It includes structuring, highlighting, and elaborating on information (Karstens and Schmitz 2023; Schukajlow et al. 2012) and supports comprehension across domains. Note-taking helps manage complex information in mathematical contexts, especially in text and numbers tasks. Organizational and elaborative strategies—like writing down key data or highlighting—can reduce errors and improve understanding (Weinstein and Mayer 1986).

This study investigates how note-taking strategies contribute to solving complex, reality-based mathematical tasks. Student performance in these tasks depends on three main factors: task difficulty (e.g., mathematical or linguistic complexity), personal characteristics (e.g., mathematical and language proficiency, cultural resources), and strategies used. While the first two have been widely studied (Knabbe et al. 2025; Leiss et al. 2024; Reinhold et al. 2020), strategies offer a direct avenue for intervention, as they are learnable regardless of prior achievement.

We examine how students’ mathematical proficiency, language competence, and cultural resources interact with their use of note-taking strategies. We propose that behaviors such as highlighting, documenting key information, and structuring processes are beneficial for solving reality-based tasks. Our focus is on the depth of information processing when employing elaborative strategies, like connecting numbers with their contextual significance, and organizational strategies, such as logically arranging information. By meticulously coding these subprocesses, this study addresses the request by Rogiers et al. (2020) for a more precise capture of strategy usage nuances.

Utilizing generalized linear mixed models, we acknowledge the hierarchical structure of our data. Our results aim to enhance understanding of how note-taking aids mathematical literacy and to guide teaching methods that effectively equip students for practical mathematical challenges.

2 Theoretical Background

Research in mathematics education has long emphasized the importance of application-oriented instruction with reality-based tasks (e.g., Blum et al. 2007; Pollak 1977) to bridge the gap between academic mathematics and real-life application (Gravemeijer et al. 2017; National Council of Teachers of Mathematics 2000; Organization for Economic Co-operation and Development 2013). However, over 20 years after the first PISA shock (Boaler 2001), the 2022 results again highlight the fragility of basic mathematical skills among secondary students (Lewalter et al. 2023). Nearly one in four students also exhibited limited reading proficiency, which poses further challenges in solving text-based, real-world math tasks.

2.1 Reality-based Tasks

Working with reality-based tasks is enshrined as an important competency in curricula (Lu and Huang 2021), and problems designed to teach this competency must include a solution-relevant reference to reality and corresponding textual explanations (Verschaffel et al. 2020).

2.1.1 Characteristics of the Task Type

While word problems describe mathematical operations that are dressed up in a text with limited real-world relevance and can usually be solved without understanding the task situation (Cummins et al. 1988; Verschaffel et al. 2000), reality-based tasks are set in actual contexts and ask complex questions out of the real-world (see Section 4: "How heavy is all the money that the thieves want to steal in the series 'Money Heist'"?). Their solution requires (re)creating the described situation mentally and translating real-world information into mathematics. This translation process—*mathematical modelling*—has been conceptualized in various frameworks and cycles (e.g., Borromeo Ferri 2007; Kaiser et al. 2011; Stillman et al. 2013). The expanded modelling cycle (see Fig. 1; Leiss et al. 2024) highlights the key cognitive steps to solve such tasks. Understanding the unique solution processes involved in reality-based tasks is critical to identifying the sources of difficulty and developing effective instructional strategies.

A task derived from the real world must be interpreted to build a mental model of the situation—central to the solution process. This model simplifies the real scenario, serving as the foundation for translation into mathematics (Leiss et al. 2024). The first stages of comprehension, simplification, structuring, and mathematizing can pose difficulties when working with reality-based tasks and lead to cognitive barriers in later phases (Clarkson 1991; Galbraith and Stillman 2006; Kaiser et al. 2011; Mayer and Hegarty 1996; Stillman et al. 2010; Wijaya et al. 2014). Especially weaker students have difficulties extracting and interrelating information from

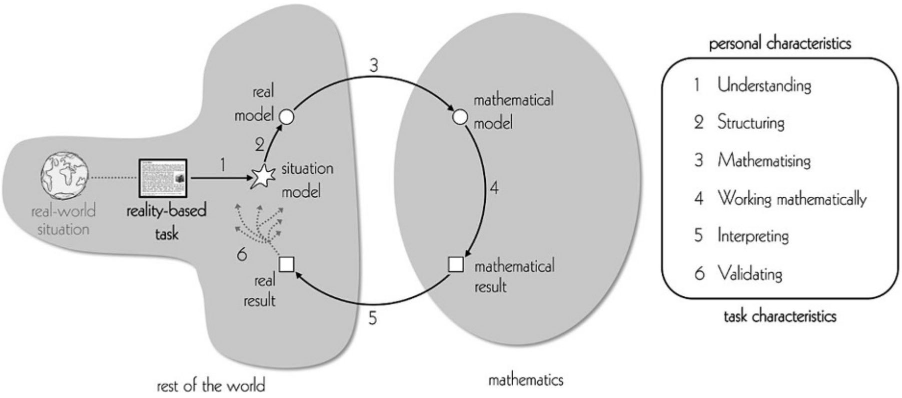


Fig. 1 Expanded modelling cycle of reality-based tasks. (Leiss et al. 2024; based on Blum and Leiss 2007)

the text (Schukajlow et al. 2012)—possibly because understanding the task requires reading and mathematics proficiency (Leiss et al. 2010; Schnotz and Dutke 2004). Creating an effective situation model and consequently resolving the task requires advanced text comprehension. Students need to grasp the intricacies of the task text upfront so that a genuine model facilitates the use of intra-mathematical concepts procedures (Daroczy et al. 2015; Niss and Blum 2020; Pongsakdi et al. 2020; Reinhold et al. 2020).

These mathematical modelling processes involve a range of reality-based tasks (e.g., Degrande et al. 2018; English and Lesh 2003; Verschaffel et al. 2020). These text-based tasks with high real-world relevance generally have lower solution rates than traditional math tasks (Daroczy et al. 2015; De Bock et al. 2003). Empirical explorations of the complex interplay of solution-relevant factors in performance situations involving reality-based tasks are rare (El Masri et al. 2017; Verschaffel et al. 2020).

2.1.2 Task Difficulty

Reality-based text tasks, which are set without photos or situational awareness, require a much longer description of the situation, involving various factors that increase the task's difficulty compared to typical tasks textbooks. The cognitive demand (Bloom 1956; Embretson and Daniel 2008; Marzano and Kendall 2007) and general requirements (Mullis et al. 2021) of the subject-specific activity have also received attention; for example, studies on the number of operators and/or numbers included in tasks show mixed results (Board 2010; Leiss et al. 2024; Sweller 1994). Cognitive load is also another factor to consider (Sweller and Chandler 1991). In traditional mathematical tasks, extraneous load is reduced by eliminating information from the text that isn't essential for the solution (Nurjanah and Retnowati 2018). In realistic tasks, the amount of information, which includes information unnecessary for the mathematical solution, can predict difficulty (Voyer 2011).

Task difficulty could also be determined by linguistic complexity, which is calculated using readability measures (Vajjala and Meurers 2012), which consider, among other things, sentence and word length. Heine et al. (2018) extended Ferrara et al.'s (2011) framework for the observed relations of linguistic elements to mathematical performance reality-based tasks by including lexis, (morpho)syntax, and information structuring for the differentiation of task texts. Leiss et al. (2019) later tested Heine et al.'s model to determine the extent of this relationship. There are significant correlations found between reading comprehension and the situation model. Research on the specific linguistic simplification of word problems reveals no positive impact on mathematical problem-solving speed (Haag et al. 2015; Walkington et al. 2019). Meta-analyses reveal contradictions in the influence of different linguistic variables (Cruz Neri and Retelsdorf 2022; Strohmaier et al. 2023).

Thus, the students' personal characteristics also predict their ability to solve reality-based tasks. Fuchs et al. (2018), Peng et al. (2020), Reinhold et al. (2020), and Strohmaier et al. (2021) identify language competence as (one of) the strongest predictors of solving reality-based tasks since text comprehension is a strong predictor for building an adequate situation model, and, thus, for solving reality-based mathematical questions (van Dijk and Kintsch 1983; Leiss et al. 2010, 2019). Adams (2003) refers to understanding through reading numbers and symbols in addition to reading and understanding words.

Also, mathematic competencies are correlated to solving reality-based tasks (Ding and Homer 2020; Plath and Leiss 2018; Prediger et al. 2018; Schilcher et al. 2021). Mathematical content-related skills and the ability to solve such tasks are also strongly correlated (Berkowitz and Stern 2018; Fuchs et al. 2006; Reinhold et al. 2020). Pongsakdi et al. (2020) found a strong relationship between arithmetic skills and the solution process.

Additionally, consider the factors influencing cultural resources (Organization for Economic Co-operation and Development 2013; Knabbe et al. 2025). These cultural resources include parental education (Piel and Schuchart 2014; Vilenius-Tuohimaa et al. 2008) or socioeconomic status (Lubienski 2000; Werning et al. 2008), which are often indicated by the number of books at home. Regarding gender, boys outperform girls in solving realistic tasks (e.g., Lindberg et al. 2010; Organization for Economic Co-operation and Development 2019). Julie (2007), Schukajlow et al. (2012), Gijsbers et al. (2020), and Vos (2020) show idiosyncrasies in how students emotionally evaluate contexts, and this positively affects the mathematics and general performance (Knabbe et al. 2025). As such, context's influence on task difficulty remains largely unexplained.

2.2 Strategies to Overcome Difficulties While Solving Reality-based Tasks

Reality-based tasks are text-heavier than traditional textbook-exercises; their solutions require enhancing comprehension and organizing information (Afflerbach et al. 2020; Karstens and Schmitz 2023), especially since the simplification of STEM texts has not yet led to better subject performance (Härtig et al. 2019). Effective real-world problem-solving demands a combination of strategies (Verschaffel et al. 2020). Such a strategy is particularly limited (Kintsch and van Dijk 1978; Paas et al. 2003; Sweller et al. 2019), and most research examines correlations between strategy application and intra-mathematical tasks (Newton 2020).

In solving reality-based tasks, the automatic decoding of words and semantic mapping of text elements requires increased cognitive resources (Bohn-Gettler and Kendeou 2014; Kintsch and van Dijk 1978; Ozuru et al. 2009; Zaitoun et al. 2023). (Meta)cognitive activities while modelling such mathematical tasks are also critical (Lester et al. 1989; Maaß 2006; Verschaffel et al. 2020). This task type necessitates a subject-specific strategy use specification (Dumas 2020).

The relationship between strategy and the learning process is rigorously investigated in education psychology (Dinsmore et al. 2020). For mathematics education, Beaudine (2022) presents a list of different strategies for mathematical reading comprehension; the categories of "Make connections across the text," "Decode text," "Take notes while reading," "Paraphrase text," and "Look for important information" are especially interesting (Beaudine 2022, p. 195).

Students are expected to adopt a strategy vis-à-vis the task (Bråten and Samuelstuen 2007; Hadwin et al. 2001); whether these strategies can be applied to reality-based tasks in mathematics education remains unclear. Cognitive comprehension strategies are highly effective in assisting students to navigate challenges in grasping and simplifying a scenario or converting a textual situation into a mathematical problem (Kintsch and Greeno 1985; Krawitz et al. 2017; Leiss et al. 2010, 2019; Vilenius-Tuohimaa et al. 2008).

Prediger and Krägeloh (2015) consider strategies in mathematical word problems. They identify typical hurdles of real-life tasks, such as the construction of the situation model and the associated individual strategies—for example, a focus on numbers and immediately start calculation (Aebli et al. 1991; Reusser 1997; Stern 1992; Verschaffel et al. 2000), a focus on signal words (Aebli et al. 1991; Daroczy et al. 2015; Mevarech et al. 2010; Nesher and Teubal 1975) and a lack of action plans and goal orientation (Littlefield and Rieser 1993). These individual strategies are usually unsustainable and require further guided strategies, such as scaffolds (Prediger and Krägeloh 2015). Some other strategies have already been the focus of prior research on mathematical modelling, having been theoretically grounded and empirically investigated in various context studies. For instance, Schmitz and Dannecker (2023) discuss formulating conclusions and hypotheses and highlighting techniques such as underlining or writing relevant information. Rellensmann et al. (2020) additionally emphasize the role of sketch knowledge. Further strategies, particularly from a cognitive psychology perspective, include using different representations, systematic trial and error, or breaking down complex problems into subproblems to use working memory (Stender 2018) effectively. Moreover, Dröse and Prediger (2018) highlight that identifying key information is a prerequisite for successful mathematical modelling. These diverse perspectives illustrate that problem-solving and modelling strategies have already been thoroughly explored, underlining the importance of systematically fostering them strategies.

2.3 Note-Taking to Externalize Covert Strategies

These strategies facilitate the externalization of cognitive processes, enhancing memory retention and offering a structured approach to addressing real-world tasks. Therefore, it makes sense to evaluate their effectiveness in predicting problem-solving and modelling outcomes performance (Leiss et al. 2019; Schukajlow et al. 2012; Wienecke et al. 2023). We distinguish between mental computation procedures and cognitive solution strategies, which are goal-oriented and flexible (e.g., Siegler 1988; Stern 1992; Threlfall 2002). While computational procedures have fixed steps, the course of action for a strategy is not fixed in advance; instead, known experience must be applied to the new situation (Stern 1992).

Note-taking is a common learning practice in school (Lahtinen et al. 1997). Pressley and Afflerbach (2012) classify it as a strategy of repetition and organization and can reduce the high demands on working memory in reality-based tasks (Leutner et al. 2009; Mikk 2008). Achieving the goal of reading—in our case, solving the math problems—brings the information to be found into focus. When these conditions are met, note-taking will enhance comprehension (Dündar 2015) and resolve hurdles to mathematical work.

In school, this occurs through highlighting text, verbatim quoting, and paraphrasing passages in notes (Karstens and Schmitz 2023; Lahtinen et al. 1997; Lonka et al. 1994; Slotte et al. 2001). In particular, different forms of highlighting and visualization of scientific texts for learning have been investigated (Leopold and Leutner 2002). Beaudine (2022) found that note-taking was used in a few tasks, but underlining and paraphrasing mathematical information was one of the most frequently used strategies.

Organization and elaboration, are next to memorization important for successful text learning (Bråten and Samuelstuen 2007; Rogiers et al. 2020; Schmitz and Dannecker 2023; Trabasso and Bouchard 2002).

2.3.1 Organization Strategies

Organization strategies can help students manage the higher general requirements (Mullis et al. 2020) and extraneous load (Nurjanah and Retnowati 2018) by allowing them to structure and prepare large amounts of information (Voyer 2011) for mathematical processing. These strategies may also overcome challenging linguistic features of the task text (Vajjala and Meurers 2012). Given that targeted simplification of word problems does not improve mathematical solution rates (Haag et al. 2015; Strohmaier et al. 2023; Walkington et al. 2019), we must investigate strategies for organizing and simplifying information that aids the solving process and, perhaps, even mask linguistic simplification through a combination of strategies in more difficult tasks (Prediger 2017).

Marginal notes and highlights are observable cognitive processes (Schmitz and Karstens 2022; Weinstein and Mayer 1986) and indicate the student's ability to comprehend reading materials (for an overview, see Chang and Ku 2015). Such student-initiated reading strategies improve academic performance (Slotte and Lonka 1999).

Writing down information is critical to mathematics education (Dündar 2015; Graham et al. 2020), since it reduces the load on working memory (Kiewra 1989; Leutner et al. 2009; Mikk 2008; Sweller 1989) by enhancing retention and the understanding of mathematical concepts (Dündar 2015; Graham et al. 2020; Hackemann et al. 2022; Kobayashi 2005; Taylor and McDonald 2007). It helps students identify solution-relevant data and represent the links and connections between each piece of information, creating a self-explanatory outline of the text (Cox 1999), which augments text comprehension (Samuelstuen and Bråten 2007). Thus, writing information as an organization and memorization strategy is expected to improve students' ability to solve reality-based tasks.

Similarly, highlighting identifies and emphasizes essential elements in the text (Azevedo et al. 2013; Bråten and Samuelstuen 2007; Schmitz and Dannecker 2023). Per Leutner et al. (2007), recognizing relevant text is an important step for further work with that text.

2.3.2 Elaboration Strategies

Elaboration strategies further build on organizational strategies for deeper comprehension and the application of knowledge to complex problems (Weinstein and Mayer 1986). Elaboration can range from a complete absence of notes to the presentation of notes as highly structured, interrelated pieces of information (Wienecke et al. 2023). Deeper processing of information ensures a more lasting memory and promotes deeper reflection. This more thoroughly processed content integrates into existing knowledge and is compared with prior knowledge. Consequently, this information can be recognized as understood, whereas information that undergoes less depth of processing is sometimes merely reproduced (Anderson 1981; Bretzing and Kulhavy 1979).

Highly rich elaboration could support the solution process (Lonka et al. 1994; Pintrich et al. 1993; Slotte et al. 2001; Staub 2006; Wienecke et al. 2023). Kiewra (1989) and Kobayashi (2005) claim that verbatim notes are ineffective unless they are used as a platform and later paraphrased and reorganized. More recently, Wienecke et al. (2023) found a link between elaboration and solving reality-based math tasks. In contrast, overt cognitive strategies, such as highlighting, paraphrasing, and elaboration, are covert or only partially visible (Rogiers et al. 2020; Rovers et al. 2019). Nevertheless, structuring may indicate a deeper level of processing (Krulik and Rudnick 1996) as it supports the elaboration of information or may even require elaboration beforehand to structure information correctly. Montague et al. (2014) especially reiterate the importance of identifying (ir)relevant information for less wordy mathematics exercises in their intervention study *Solve it!*. In mathematics classes at German primary schools, students are taught to use structuring aids such as "given and sought" to actively think about the data provided and the question (Buschmeier 2017; Dröse and Prediger 2018). This involves repetition and content structuring (Leutner et al. 2009). Since data organization and structuring are particularly important when dealing with non-routine complex problems (Krulik and Rudnick 1996), organizing the different phases of the modelling cycle can have a supportive effect (Verschaffel et al. 2020). Such notes are created through high-level, in-depth information processing and are often meant to be self-explanatory (Rittle-Johnson et al. 2017).

2.3.3 Influencing Factors On Note-Taking

Much of the research on note-taking has focused on learning notes—i.e., the classic way of taking notes during lectures or when reading texts. The emphasis here is on memory performance, comprehension, and exam preparation. Therefore, the question of which individual prerequisites promote taking solution-supporting notes remains largely unanswered.

Language Skills: Students with higher linguistic competence produce more structured and relevant notes than those with lower language skills comprehension (Reddington et al. 2015). While there were hardly any differences in performance among men depending on their linguistic competence, the quality of the notes varied significantly more among women (Reddington et al. 2015). Writing speed plays a subordinate but not insignificant role here: while some studies found a positive correlation between handwriting speed and the quality of notes (Manzi et al. 2017), others were unable to demonstrate any significant effects (Reddington et al. 2015).

Mathematical Skills: Studies show a correlation between learning success in mathematics and the quality of notes—specifically, whether notes are taken at all and how they are organized. Not only is the content of the notes relevant, but so is their formal design (e.g., use of color, structure, and legibility) (Dündar 2015). This study also predicts that successful students in mathematics tend to take contextual notes, use different forms of presentation, and supplement their notes with content from other sources. This indicates that taking thoughtful notes is closely linked to mathematical understanding (Dündar 2015).

Gender: On average, female learners take notes more frequently and of higher quality than male learners. This is reflected in the amount of information noted and its structure and relevance (Reddington et al. 2015). Wienecke et al. (2023) identified note-taking as a solution strategy, discovering a link between being female and higher note-taking frequency. Similarly, Slotte et al. (2001) substantiated these findings, indicating that female participants engaged in open learning strategies like note-taking more often than their male counterparts participants.

Cultural Resources: There seems to be little research on cultural resources, such as the educational background of parents or the "books at home" variable used several times in PISA (Mang et al. 2018), but Wienecke et al. (2023) documented a relationship between higher socio-economic status and note-taking behavior.

The combined theoretical and empirical evidence emphasizes the essential influence of language and strategy application in addressing mathematical tasks grounded in reality. Although numerous studies have shown the effects of language proficiency, mathematical ability, and personal or cultural background characteristics on performance, the distinct impact of note-taking strategies in this context remains underexplored. Earlier research mainly concentrated on broad strategy application, frequently overlooking the intricate interactions among task difficulty, personal resources, and visible strategies throughout the solution process.

3 Research Questions and Hypotheses

To address this gap, as part of Project VAMPS[1], the present study aims to systematically investigate how different note-taking strategies and their quality manifest during task processing, how they are influenced by personal and task-related characteristics, and to what extent they contribute to a successful solution process. Accordingly, we summarize our three research questions:

RQ1: To what extent are the strategies of (a) highlighting, (b) writing, (c) evaluating the relevance of information, and (d) elaborating information applied with varying quality in students' notes during the solution process of reality-based tasks at different language levels?

RQ2: Which personal and task-related language characteristics are associated with creating notes and their quality to derive the solution in reality-based tasks?

RQ3: Is there a connection between correctly solving approached reality-based tasks and the qualitatively assessed application of (a) highlighting, (b) writing, (c) evaluating the relevance of information, and (d) elaborating information?

Owing to the research desideratum described above, no clear expectations can be formulated for the particular type of task selected here. However, some hypotheses can be derived from theory that is transferable to these tasks.

For RQ1, we hypothesize that students will exhibit all types of note-taking during the solution process of reality-based tasks, and each strategy will vary based on the linguistic complexity and real-world context of the task since effective solution process of reality-based tasks demands a combination of strategies (Strohmaier et al. 2021; Verschaffel et al. 2020). For RQ2, we hypothesize that personal characteristics, such as mathematics and language proficiency (Dündar 2015) and cultural resources (Wienecke et al. 2023), will predict the type and quality of notes students take. We do not assume any hypotheses regarding gender effects on note-taking as a learning strategy because the available results are inconclusive (e.g., Slotte et al. 2001; Wienecke et al. 2023). For RQ3, we hypothesize that, in addition to the empirically established influencing factors of language proficiency (independent variable; Strohmaier et al. 2021), mathematics proficiency (independent variable; Ding and Homer 2020; Strohmaier et al. 2021), cultural resources (independent variable; Knabbe et al. 2025) and task-difficulty (independent variable; Heine et al.

[1] Variation Tasks Mathematics Physics Language, founded by the German Research Foundation, project No. 417017613.

2018; Klotz et al. 2025), note-taking (independent variable; Leiss et al. 2010) is positively related to the solution process (dependent variable).

4 Methods

4.1 Sample

Our convenience sample consisted of 428 students from a comprehensive school that includes all academic tracks. Only students without missing data for the relevant variables were included in the analysis, resulting in a final sample of 395 students. Each student performed three distinct reality-based tasks, leading to a total of 1064 completed tasks. The distribution of students by grade level was as follows: 29.73% in 7th grade, 73% in 8th grade, 28.35% in 9th grade, and 16.96% in 10th grade, with an average age of 14.85 years (SD = 1.26). Among the participants, 50.13% identified as girls, 47.30% as boys, and 2.57% did not identify as either. Furthermore, 86.93% of the students had no immigrant background, while 9.52% spoke a language other than German at home, which aligns closely with typical demographic trends region (Mang et al. 2018). The study was conducted in accordance with all applicable ethical guidelines. Approval was obtained from both the State School Authority and the Ethics Committee of Leuphana University. In addition, written informed consent was collected from the legal guardians of all participating students. This ensured that participation was voluntary and based on informed consent.

4.2 Procedure & Instruments

Trained test administrators administered a paper-and-pencil test on a single school day across 26 classes. To avoid class effects, 12 different test booklet versions were randomly distributed among the participating classes. After a brief test instruction, each student had 50 min to individually complete one of the 12 booklets. Each booklet consisted of the following four parts:

Part 1: Language proficiency test (cloze format) (5 min)
Part 2: Mathematical reality-based tasks with a text comprehension questionnaire (30 min)
Part 3: Mathematics proficiency test (computational test to functional relationships) (10 min)
Part 4: Questionnaire personal characteristics (5 min)

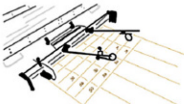
4.2.1 Mathematical Reality-based Tasks

The primary tool for the survey addressing all research questions is the reality-based tasks. We utilized these tasks to document solution paths along with notes and, inevitably, the solutions themselves. The research project consisted of nine reality-based tasks crafted by trained mathematicians to enhance modelling literacy. Each task illustrated one of the three real-world contexts, tailored to various language lev-

Fig. 2 Description of the real-world context of the *Task MH* on *LL2*

Money Heist

In the series „Money Heist", 8 criminals are preparing a major burglary in which they want to rob the banknote printing plant in Spain. However, they are not after the money in the bank's vault, so the robbery is a little special. Instead, the criminals want to get into the bank's printing plant to print money themselves. They plan to produce a total of 2,400,000,000 euros, with which they want to make off. The criminals want to limit themselves to producing 50 euro bills, as these are not as conspicuous when paying as larger bills. However, production is relatively complex, as special banknote paper and ink are required. In this way, 7,500 sheets of money paper can be printed in one hour, which requires 10 kg of special ink. Each of these sheets of paper has space for 8 rows, each with 5 columns of 50 euro bills, which are then cut into individual 50 euro bills by a machine. Each 50 euro bill has a length of exactly 140 mm and a width of 77 mm, while the weight is exactly 0.92 g per bill. After printing, 10,000 of the 50 euro bills are always packed together in a transparent plastic bag, which are then all stacked on top of each other. At the end of the series, the criminals escape from the banknote printing plant by disguising themselves. They manage to get hold of a large number of these plastic bags, each of which is worth € 500,000.



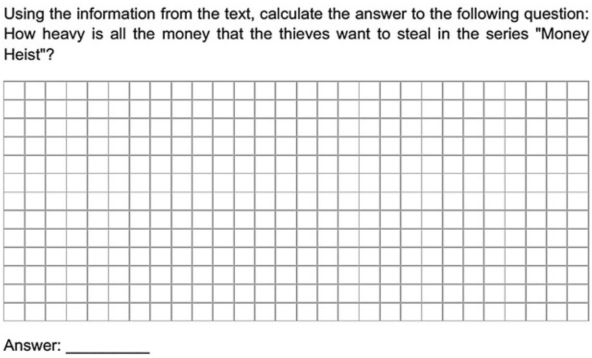
els and mathematical concepts question. The real-world contexts included YouTube [*Task YT*], the Spanish television series *Money Heist* [*Task MH*] (see Fig. 2), and Letterpress [*Task LP*]. Based on real-world problem situations, the description was not specifically created for testing purposes.

The factual texts in the tasks contained an average of 212.4 words ($SD = 19.7$) and numerical and textual information. That is the mathematically relevant information was not limited to a short text passage but was spread over the entire body of the text, which made reading and understanding the entire text a prerequisite for solving the mathematical question.

Compared to classic modelling tasks, such as Kaiser (2017), the students must assess the relevance of various information, but they do not have to make any assumptions. For reasons of test economy, all the figures needed to answer the question are provided in the text, which is why our tasks do not align with reality-based tasks (Verschaffel et al. 2020) as pure modelling tasks. Nevertheless, to solve these tasks, students need the ability to extract information from real contexts and classify it mathematically, work with it, and compare the results with reality at the end again (Verschaffel et al. 2020). To signal that these items are not merely dressed-up word problems, nor are they clearly modelling tasks as Kaiser defines them, we refer to them as reality-based tasks.

Linguistic variations can predict solving real-life tasks and apply strategies (Haag et al. 2015; Johnson and Monroe 2004; Walkington et al. 2018). Thus, even if the effects are likely to be small (Strohmaier et al. 2023), a team of linguists used the original text created by the mathematicians to create three versions of the text with different linguistic characteristics (Hackemann et al. 2022; Heine et al. 2018).

The three different reality-based tasks respectively exhibited three levels of complexity of academic language features (LL1, LL2, and LL3, where "LL" indicates the "language level"). The LLs rotated randomly so that each test booklet contained one task with LL1, one with LL2, and one with LL3 in any order and for any task context. The text was first divided into discrete units of information that cor-


Using the information from the text, calculate the answer to the following question: How heavy is all the money that the thieves want to steal in the series "Money Heist"?

Answer: _________

Fig. 3 Mathematical question for *Task MH*

responded to the meaning expressed in the original text. Then, a version close to the original text was created (*LL2*), which adhered to rigid construction principles to maintain consistent structural elements across the tasks and later served as the foundation for developing two other versions of the text: one with fewer features of academic German (*LL1*) and one with more (*LL3*) than *LL2*. The main clause/subclause ratio (syntax) was an important construction factor. Thus, *LL1* primarily used main clauses, while *LL2* consisted of a sequence of main and subordinate clauses. *LL3* only utilized syntactic structures that were more complex than those in *LL2* and typically exchanged subordinate clauses for participles. The second key point of variation was the verb forms. LL1 predominantly used verb forms instantiated in a single word, whereas *LL2* employed verb forms made up of several elements. In *LL3*, verb forms consisting of three elements were also utilized. Semantic transparency, verbal/nominal style, deixis, and impersonal expressions are other aspects that differ in the language levels. We gradually increased the linguistic structural complexity and reduced semantic transparency to complicate the decoding process (McNamara and Magliano 2009) and the overall cognitive load during decoding (Mikk 2008). We expected students to use organizing strategies to overcome the high text complexity at *LL3*.

The situation description was followed by a question that required a mathematical solution from the content area of mathematical functions (Fig. 3), ensuring that all students, regardless of their ability level, had covered the content in class[2]. A preliminary study showed that the students had a similar amount of and relatively high prior knowledge of the task context (*Context YT*: 3.21, *Context MH*: 3.17, *Context LP*: 2.94), which they had to rate themselves on a four-point Likert scale. Similarly, students exhibited a rather higher level of interest (*Context YT*: 3.07, *Context MH*: 3.43, *Context LP*: 2.78) on a four-point Likert scale (Knabbe et al. 2025).

The mathematical questions were crafted using a coding manual that maintained a consistent complexity level, encompassing the necessary numbers for solutions

[2] The shift in instructional content resulting from the COVID-19 pandemic caused some intra-mathematical difficulties in the computational test; however, these problems did not affect the reality-based tasks nor the notes.

and the requirement level aligned with the German curriculum fluency (Leiss et al. 2024).

Despite ensuring maximum task authenticity, differences between the tasks became apparent, such as in the order of the required information in the text. In *Context YT* and *Context LP*, for example, the numbers did not appear in the order they would logically appear in the calculation. The units in *Context LP* were also more complex than those in *Context YT* and *Context MH* because they were calculated in euros per hundred pages. The complexity of the calculation also differed. *Context LP* necessitated taking an intermediate step, which students were expected to solve individually, and the results would be used to continue the calculation. The students were instructed to first calculate the price of a book before determining the total price. In Context MH, the students were directed to find the three numbers necessary for the solution, which totaled 12. Highlighting or writing these three numbers after reading the question would be the most effective strategy for this task. Here, it was already possible to distinguish whether the numbers had been written out without additions, arranged in their own order, or even presented with additional information. These notes prior to the actual calculation helped students form adequate situational, real, and mathematical models that clarified the relationships between the numbers before the calculation began.

4.2.2 Task-related Text Comprehension

To study the prediction of linguistic variation on text comprehension, we developed a text comprehension instrument as part of the survey (Klotz et al. 2025). The students solved the comprehension questions immediately after reading the context and prior to addressing the task. We posed questions regarding the essential information necessary for understanding the task and constructing the situation model. All reality-based tasks were examined for the mathematically essential information they included, and nine single-choice comprehension questions (e.g., "The criminals manage to escape with €2.4 billion") were created. Each question could be answered with the responses "correct," "incorrect," or "statement cannot be assessed with the information from the text" (the correct response was coded 1, and the incorrect/missing response coded 0) ($\alpha = 0.627$). In this manner, we assessed the prediction of note-taking under the control of text comprehension as a significant predictor of solution, consistent with extant findings.

4.2.3 Computational Test

In addition to the task-specific characteristics, personal characteristics were collected to answer RQ2 and as further predictors to specify RQ3. The intra-mathematical skills were measured by an adjusted subject test adapted from DEMAT 9[3]. Given its comparability to reality-based tasks, the content area of the tasks consisted of

[3] The DEMAT 9 is a standardized test to assess the mathematical abilities of students within the described sample that has already been tested several times. The internal consistency of the curricular valid tasks of the DEMAT 9 is $r = 0.94$ (Schmidt et al. 2012).

functions. All students completed six tasks that required them to solve equations ($5 + x = 3, x = ?$, see Schmidt et al. 2012). To assess mathematics proficiency, a sum score was calculated from the correctly completed tasks ($M = 0.373$, $SD = 0.316$, $\alpha = 0.782$).

4.2.4 Language Proficiency

A language cloze test was administered to test general language skills (DCLL +3[4]). Cloze tests are a reliable and valid measure of general language proficiency (Grotjahn et al. 2002; Grotjahn and Drackert 2020). The score of the cloze test is based on the proficiency with which students correctly and individually identify the searched word for each gap (Harsch and Hartig 2016). The response's orthographic and grammatical accuracy is assessed according to the case. Overall, the text comprised 96 words. Students were given 5 min to fill 25 gaps. The language proficiency score depended on the number of correctly spelled words, with a range from 0–1 ($M = 0.86$, $SD = 0.161$, $\alpha = 0.87$).

4.2.5 Socio-demographic Characteristics

The socio-demographic questionnaire included questions on gender (male = 1; female = 2; non-binary = 3), age, and social background. In addition to the year and country of birth, we included the birth countries of the students' parents to indicate any migration background, as well as the first-learned language and primary language at home.

For information on cultural resources, we used the PISA-Scale. Students selected the number of books at home on a six-point Likert scale ranging from "0–10 books" to "more than 500 books" dummy coded as 1 and 6, respectively (Mang et al. 2018). As in PISA 2018, reading enjoyment, pleasure, and strategies were questioned. Most interestingly, reading strategies were indicated on a four-point Likert scale, ranging from "never or almost never" to "almost always," and included questions on whether the participant marked passages in texts or took notes (adapted from Wagner et al. 2011).

4.3 Coding

Following the distinction between Organization and Elaboration, we focused on the two common organizational strategies for pointing out the required information to solve the task (Highlighting and Writing out, for example, in Schmitz and Dannecker 2023). We summited all elaborating mathematical notes since missing concepts to differentiate between different forms (see Table 1).

To distinguish between mental computation procedures and cognitive solution strategies (Siegler 1988; Stern 1992; Threlfall 2002), we only consider as "notes" those writings that relate to the formation of the situation model and real model or

[4] The DCLL +3 is a standardized test for grades 7 and 8, and is available from the Test Development and Diagnostics Unit of the Hamburg Institute for Educational Monitoring and Quality Assurance.

Table 1 Summary of the coding of the metacodes

	Description	Conditions for Score 1	Conditions for Score 0	Code in Student Solution from Fig. 4 with explanation
Response	We defined a correct solution as one that included both the numerical result and the correct unit (e.g., 44,160,000 g, 44.16 t, or 44,160 kg), assigning a score of 1 to such responses to ensure clarity in the analysis	Students who have the correct numerical answer with the correct unit	Student with either a wrong number or/and a wrong unit	The student got the wrong number but the correct unit—thus the student got a score 0 for our response variable
Highlighting all required information	The *highlighted* text passages were individually coded so that each piece of information was coded according to whether it was underlined, highlighted, marked, circled, etc. Since the mathematical question targets specific information, some details are more important for the solution process. Thus, each piece of information was categorized as required or irrelevant in advance. In this code, we show the case where a student has underlined all the required information	Students who highlighted all required information (€ 2.4 billion, € 50 bills, and 0.92 g for *Task MH*)	Students who only highlighted one, two, or none of the required pieces of information	Since we do not have a student's work here, we cannot evaluate this code. In this example, the student received a code of 0 because he did not underline "€ 50 bills."
Writing all required information	The required information *written out* of the text was coded in detail. Each piece of information in the task text was individually coded as written out or not regardless of whether this information was required to solve the mathematical task. Thus, there were individual codes for all information as the base for a code, which summarizes the number of the relevant information written out and the irrelevant information written out. To assess the prediction of the strategy used, we constructed a code that measured the optimum cases	Students who had written out all required information to solve the task (e.g., € 2.4 billion, € 50 bills, and 0.92 g for *Task MH*)	Students who only wrote out two, just one, or none of the required pieces of information	In this example, the student received a score of 0 because the information "€ 50 bills" was information
Taking irrelevant notes	Irrelevant notes included all information that was clearly not required for the mathematical solution, and included *further sketches* and *remarks that do not support the solution process* (e.g., "eight criminals want to steal the money") or *statements that the task could not be solved* or the inability of the student to understand the task. We did not include students who highlighted irrelevant information, since highlighting also functions as a reading strategy	Students who had presented or explained something that was not required to solve the question	Students who only have relevant and helpful information in their notes	In this example, the student did not write any irrelevant information, and, therefore, received a score of 0

Table 1 (Continued)

	Description	Conditions for Score 1	Conditions for Score 0	Code in Student Solution from Fig. 4 with explanation
Taking elaborative notes	In addition to the organizational strategies of highlighting and writing, we observed students' *elaboration* of information, that is, their ability to perform in-depth processing of the content to present rich notes. We coded the extent to which *relationships* were established between numbers and their meaning, along with helpful *sketches*. We observed the *order of the information* written out to see how it differs from the order in which it appears in the text; this helped us assess whether the students had merely copied information or used their own logic before continuing to work with it. Also, attention was paid on the structure of the checkered field: *Structuring Features* in mathematics can structure the calculation as well as the notes (e.g., to structure and clarify the given information, determine which quantity is being searched for in the calculation process, or mark a result)	Students notes which shows for at least one of the categories printed in italics received a score of 1; e.g., a student who has placed the information into a logical order of his or her own	Students who merely highlighted or wrote out information in the order of the text or did nothing at all	In this example, the student provided an example of a representation of relationships. The student wrote that 0.92 g was the weight of "one banknote" and that the €2.4 billion was "what they wanna steal." The student had structured the "given" information and his "calculation path." The student received a score of 1 for his ability to process the information further

It was actively decided to use only the "all required information" codes, although each piece of information was coded individually, as these guarantee not only the choice of strategy, but also the correct application and the quality of this strategy

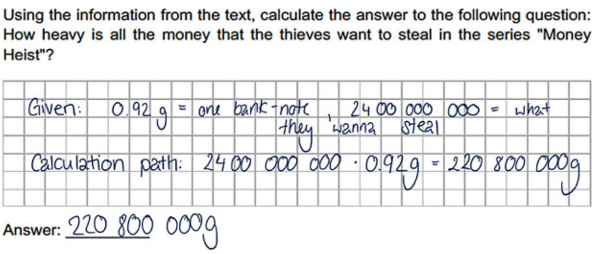


Fig. 4 Original student solution translated into English by the authors

the features used to structure the solution process for the task. The mere calculation path, representing the mathematical further processing, is not coded as "notes."

4.4 Analysis

Data were analyzed using SPSS-version-28.0.0.0(190) and R-4.3.3 statistical-software. To answer RQ1, we calculated descriptive statistics on the quality of the different types of notes in the three tasks. To answer RQ2, we observed task-specific correlations between various note-taking characteristics and their quality. The correlations were corrected for multiple comparisons using the Benjamini–Hochberg procedure to control the false discovery rate (Benjamini and Hochberg 1995). For RQ3, generalized linear mixed models (glmer function in package lme4; Bates et al. 2015; Boeck et al. 2011) were computed with personal and note-taking characteristics as independent variables and correct task solution as the dependent variable. Generalized linear mixed models are preferable for our planned analysis because they account for random effects and control for interfering factors while enabling the modelling of complex, hierarchical data structures (Burnham and Anderson 2004; Nakagawa et al. 2017). With multivariate analyses and flexibility in the distribution of variables, they offer improved predictions compared with correlations. We used a significance level of $\alpha = 0.05$ for all analyses.

5 Results

5.1 Frequency Analyses—RQ1

Regarding RQ1 (*To what extent are the strategies of (a) highlighting, (b) writing, (c) evaluating the relevance of information, and (d) elaborating information applied with varying quality in students' notes during the solution process of reality-based tasks at different language levels?*), we distinguished between *Task YT, Task MH*, and *Task LP* and *LL1*, *LL2*, and *LL3* in the frequency analysis. The corresponding solution rates were added to Fig. 5 to aid in interpreting the results. The table displays the number of cases in the first row, the solution rate, and the percentages of the focused note-taking characteristics in the second row. Significant differences in ANOVA and Tukey HSD are also emphasized by connecting lines and significance asterisks.

	Descriptive analysis of the different types of notes						
	Task YT	*Task MH*	*Task LP*	*LL1*	*LL2*	*LL3*	Total mean
Number of cases *n*	327	374	363	363	355	346	1046
Solution rate *M*	0.07	0.39	0.27	0.25	0.26	0.25	0.25
Highlighting all required information	3.06%	2.67%	6.89%	5.79%	3.67%	3.18%	4.23%
	*** (Task YT–Task MH)	*** (Task MH–Task LP)					
Writing all required information	6.12%	0.27%	6.06%	3.58%	4.51%	4.04%	4.04%
	*** (Task YT–Task LP); *** (Task YT–Task MH)						
Taking irrelevant notes	12.23%	1.87%	4.13%	7.44%	5.35%	4.62%	5.83%
	*** (Task YT–Task MH)	*** (Task MH–Task LP)			** (LL2–LL3)		
Taking elaborative notes	57.49%	40.64%	61.70%	53.17%	57.75%	47.98%	53.00%

Fig. 5 Frequency analyses of different note-taking characteristics by task in percentage

The frequency distributions of the correct application of different note-taking strategies reveal insightful trends across the tasks and their language level (see Fig. 5). *Task LP* had the highest percentage of students highlighting all required information, likely owing to the complexity of the calculations and units involved. This task also had a notable proportion of relevant written notes (6.06%, a significant difference to *Task MH* with analysis of variance and Tukey's range test[5]) and the highest percentage of elaborative notes (61.70%), indicating that students were actively engaged with the material, despite 4.13% of the solutions including irrelevant notes.

Task MH showed the least amount of writing; only 0.27% (significant difference to both other tasks) of the students wrote out all the relevant information to solve the task, similar to highlighting at 2.67%. This task had the fewest elaborative notes and notes in general, possibly because of its high solution rate, suggesting that students felt confident in their ability to solve the problems without additional notes.

Task YT had the highest percentage of irrelevant notes at 12.23% (a significant difference to both other tasks), while *Task MH* had the fewest (1.87%). *Task YT* had the highest percentage of solutions with all required information written out. Overall, the tendencies of making notes seemed to depend on the difficulty of the task context and the required operations: *Task MH* had the lowest level of elaborative note-taking (40.64%, a significant difference to both other tasks), indicating that students avoided making deeper connections or expanding upon the information in their notes. Compared with Task LP, we considered a tendency for more superficial note-taking behaviors in this task, where elaborative strategies were more prevalent.

On average, the students took fewer notes while solving tasks on a higher language level. The proportion of students who highlighted all the required information decreased with increasing language level, from 5.79% at *LL1* to 3.18% at *LL3*. Possibly, more complex texts make it more difficult to identify relevant information.

[5] To further investigate the significant differences between the groups, a Tukey HSD (Honestly Significant Difference) test was conducted following the one-way analysis of variance (ANOVA). The Tukey test allows for pairwise comparisons of means while controlling for multiple testing to reduce the risk of Type I errors (Nanda et al. 2021). Significant differences are indicated by *p*-values less than 0.05.

LL2 (57.75%) had the most elaborate notes compared with *LL1* (53.17%) and *LL3* (47.98%). Elaboration is also where the only significant difference between *LL2* and *LL3* becomes apparent. All other differences between the frequencies of note-taking on the different language levels were insignificant.

5.2 Relations to Personal and Task Characteristics—RQ2

Regarding RQ2 (*Which personal and task-related language characteristics are associated with creating notes and their quality to derive the solution in reality-based tasks?),* we calculated the relationship with a Pearson correlation because of the metric personal characteristics. In addition to the correlations, Table 2 presents the minimum and maximum values of the scale, the mean, and the standard deviation. Here, we observe the direct correlation between the different types of notes (vertical) and personal characteristics (horizontal).

In summary, the correlations show that personal characteristics are strongly linked to students' note-taking methods and their cognitive involvement with the content. Greater language proficiency correlates with a higher inclination to capture all essential information (language proficiency: $r = 0.08$**). Higher mathematics proficiency was also associated with fewer irrelevant notes ($r = -0.08$*). As expected (Slotte et al. 2001; Wienecke et al. 2023), girls took more high-quality notes than boys ($r = 0.09$**; $r = 0.11$**; $r = 0.08$*) in the reality-based mathematical tasks.

Cultural resources significantly correlated with supportive highlighting ($r = 0.11$**) and elaborating ($r = 0.09$*). Self-reported strategy usage indicated that the claim of applying strategies correlated to the organization strategy of highlighting ($r = 0.16$**), while text comprehension was significantly correlated to writing ($r = 0.09$**) and taking elaborative notes ($r = 0.08$*). Interestingly, the appearance of irrelevant notes does not correlate significantly.

5.3 Effect On Success of Solution—RQ3

Regarding RQ3 (*Is there a connection between correctly solving approached reality-based tasks and the qualitatively assessed application of (a) highlighting, (b) writing, (c) evaluating the relevance of information, and (d) elaborating information?*), we decided to go beyond simply analyzing the quality of the different types of notes and their relationship to personal. Understanding the influence of personal characteristics and task complexity on note-taking is crucial, but the key question is whether these notes facilitate effective task-solving. Consequently, we aimed to identify which types of notes are oriented toward solutions and genuinely aid in task performance. To achieve this, we analyzed the relationships between the observed note-taking practices and successfully obtained the correct answer solution (see Table 3).

Notably, there is a positive correlation between highlighting ($r = 0.07$*) and writing out ($r = 0.09$**) all required information, indicating that students who successfully identified and recorded key information were more likely to reach the correct solution. The distinction between the tasks is most expressed in that writing is significant for *Task MH* and LL3 ($r = 0.26$*; $r = 0.16$**) and none of the others. The absence of irrelevant notes is also associated with the correct solution ($r = -0.12$**),

Table 2 Correlations with personal characteristics

	Language proficiency	Mathematics proficiency	Gender (0: = male, 1: = female, 2: = divers)	Cultural resources	Self-reported strategy usage	Text comprehension
	M (SD) 19.23 (4.53)	M (SD) 0.40 (0.32)	M (SD) 1.56 (0.55)	M (SD) 3.41 (1.21)	M (SD) 2.10 (0.56)	M (SD) 0.61 (0.19)
	Min: 0 Max: 25	Min: 0 Max: 1	Min: 0 Max: 2	Min: 1 Max: 5	Min: 1 Max: 4	Min: 0 Max: 1
Highlighting all required information	0.05	0.06	0.09*	0.11**	0.16**	0.05
Writing all required information	0.08*	0.07	0.11**	0.05	0.07	0.09*
Taking irrelevant notes	–0.06	–0.06	–0.03	–0.00	0.02	–0.06
Taking elaborative notes	0.09**	0.06	0.08*	0.09*	0.05	0.08*

All correlations were checked using the Benjamin–Hochberg method and the new *p* values were added (Benjamini and Hochberg 1995)
M and *SD* represent the mean and standard deviation, respectively
The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming 2013)
* indicates $p < 0.05$. ** indicates $p < 0.01$

Table 3 Correlations with the correct solution

	Correct solution						
	Task YT	*Task MH*	*Task LP*	*LL1*	*LL2*	*LL3*	Total
Note-taking							
Highlighting all required information	0.02	0.11	0.08	0.05	0.06	0.13	0.07*
Writing all required information	0.03	0.06	0.26**	0.02	0.09	0.16*	0.09**
Taking irrelevant notes	–0.07	–0.03	–0.13*	–0.12	–0.11	–0.14	–0.12**
Taking elaborative notes	0.21**	–0.09	0.39**	0.10	0.07	0.11	0.09**
Personal Characteristics							
Language proficiency	0.07	0.39**	0.32**	0.29**	0.21**	0.33**	0.28**
Mathematics proficiency	0.09	0.38**	0.31**	0.24**	0.24**	0.31**	0.26**
Gender	–0.11	–0.01	0.05	0.01	–0.04	0.00	–0.00
Cultural resources	0.14*	0.13*	0.12	0.12*	0.04	0.18*	0.11**
Self-reported strategy usage	–0.05	0.10	0.07	0.08	–0.02	0.09	0.06
Task-related text comprehension	0.18**	0.37**	0.33**	0.31**	0.32**	0.25**	0.29**

All correlations were checked using the Benjamin–Hochberg method and the new p values were added (Benjamini and Hochberg 1995)
M and *SD* represent the mean and standard deviation, respectively
The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming 2013)
* indicates $p < 0.05$, ** indicates $p < 0.01$

confirming the need to differentiate between the relevance of information. Regarding the preparation of elaborative notes, we find positive correlations with the moderately difficult ($r = 0.39$*, *Task LP*) and difficult ($r = 0.21$*, *Task YT*) context. Overall, elaborative notes reveal a positive correlation ($r = 0.09$**), suggesting that elaborated notes are related to better performance.

Regarding personal characteristics, both language proficiency ($r = 0.28$**) and mathematical proficiency ($r = 0.26$**) proficiency are, as expected, positively correlated with reaching the correct solution. However, in *Task YT*, the most difficult task, neither language nor mathematical proficiency shows a correlation, which we attribute to the low solution rate. Gender does not reveal itself as significant in any task after correcting the p values through the Benjamini–Hochberg method (Benjamini and Hochberg 1995), while cultural resources show a positive correlation ($r = 0.11$**). Finally, text comprehension correlates ($r = 0.29$**) with achieving the correct solution.

Correlations capture only the linear relationships between variables and do not account for potential confounding factors or the hierarchical structure of the data. A generalized linear mixed model allows for a more comprehensive analysis by considering both fixed and random effects, providing a better understanding of the influence of variables in complex datasets. The models presented here focus on predicting the correct solution for all tasks. All predictors that demonstrated a significant correlation with the correct solution were selected for the models. To calculate the generalized linear mixed model, the independent variables were standardized to ensure better comparability of the coefficients. To analyze the binary outcome variable (0 = incorrect solution, 1 = correct solution), a generalized linear mixed-effects model (GLMM) with a logistic link function was estimated (Bates et al. 2015). Random intercepts for tasks and students were included to account for the hierarchical structure of the data[6].

All models show a significant positive correlation between mathematics proficiency and the probability of a correct solution. The effect remains relatively constant in all models (e.g., $\beta = 0.41$–0.51, all with $p < 0.001$). As expected, higher mathematical proficiency has a consistently strong positive observed relation to the solution rate. Our finding for language proficiency is similar across all models (e.g., $\beta = 0.57$–0.80, all $p < 0.001$). However, cultural resources do not predict the correct solution when both proficiencies are controlled for. Although overall models no significant effects for the factor ‘cultural resources’ were found in our model, the post-

6 Including classroom membership as a random effect was not feasible because of unbalanced group sizes, which resulted in model singularity and convergence issues. Additionally, owing to the huge dataset, the students were integrated as random effects. Also, grouping by age would have resulted in only four groups, which is considered too few for robust estimation of stable random effects (Lin and Chen 2015). The task contexts (*YT, MH, LP*) were not explicitly modeled as a random effect. However, they are implicitly captured by including the nine different tasks, representing combinations of context and language level (e.g., *YT* at LL 1, *MH* at LL 1, etc.) and modeled as a random intercept.

We considered the inclusion of random slopes for the predictors, but models with random slopes frequently resulted in singular fits or convergence issues, likely owing to the small number of tasks ($n = 9$) or insufficient observations per group. Therefore, we opted for a more parsimonious model with random intercepts only to ensure model stability and interpretability. Future studies with larger sample sizes per grouping level may benefit from more complex random-effects structures.

hoc power analysis[7] (1000 trials, small effect) indicates a low detection probability (8.7%), suggesting that the current sample size may be insufficient to reliably detect smaller to moderate effects (Green and MacLeod 2016).

Model 2 included the task-related variables, but the constructed LL remained insignificant, with $p > 0.05$, whereas text comprehension was positive and significant (β= 0.60***). Again, understanding the text is crucial for the correct solution. This model also has a significantly higher explanatory rate (marginal R^2 significant chi-square test $p < 0.001$).

In Model 3, the measured quality of note-taking strategies were added. Note also that the first organizational strategy, highlighting all relevant information, is not significantly related to the task's solution rate. However, the second organization strategy, writing all required information, does (β= 0.23**). Avoiding irrelevant notes also affects reaching the correct solution (β= –0.35*). As expected, taking elaborative notes has a strong positive effect as well (β= 0.37***).

Model 4 reveals a significant positive interaction between text comprehension and elaborated notes. Regarding model fit, Model 4 outperforms all other comparisons, exhibiting the lowest AIC (936.10) and deviance (908.10) values, along with the highest explained variance (Marginal $R^2 = 0.288$; Conditional $R^2 = 0.560$).

Overall, it can be stated that both making elaborate notes and text comprehension predict the correct solution independently. Despite this, an interaction between these two predictors can be assumed. People with a higher level of text comprehension may process information better and think more deeply. If these individuals now take elaborate notes, the positive effect of text comprehension on solution competence could strengthen, as they can better extract the relevant information and process it in their notes. Similarly, elaborate notes require recording information and adding context, links, or explanations. Such notes necessitate a certain level of in-depth understanding of the text. This means someone who understands the text better can take elaborate notes, greatly aiding task completion. For both groups (students with low text comprehension in red and high text comprehension in blue), taking elaborate notes leads to a higher probability of a successful solution. However, students with a high level of text comprehension benefit more from taking elaborate notes than those with a lower level of text comprehension, as evidenced by the steeper slope of the blue curve.

To assess the interpretability of the non-significant interaction effect between highlighting and writing out relevant information ($p = 0.052$), we conducted a post-hoc sensitivity power analysis using the R package simr (Green and MacLeod 2016). The resulting power was 0%, indicating limited sensitivity to detect such an effect with the current sample size (395 students across nine tasks). This result suggests that the lack of significance for this interaction could be because of insufficient statistical power rather than a true absence of the effect.

[7] We conducted post-hoc power analyses using the R package *simr* (Green and MacLeod 2016) to evaluate the sensitivity of our models to detect small effects. The function powerSim() allows for the estimation of statistical power based on simulated data derived from the fitted mixed-effects models. This approach is particularly suitable for complex hierarchical designs and offers insight into the likelihood of detecting effects of a given size with the current sample. The results of these simulations help interpret non-significant findings in light of the model's statistical power.

Overall, the last model with the note-taking predictors has the highest explained variance (marginal $R^2 = 0.288$ and conditional $R^2 = 0.560$, see Nakagawa et al. 2017) and has a significantly better model fit (Chi-square test $p < 0.001$, Deviance = 908.096, Akaike information criterion = 936.096, see Burnham and Anderson 2004). Controlling for the strong predictors of math and language proficiency and text comprehension, note-taking, and especially writing all required information and elaborative note-taking, which demands a higher depth of processing, has a significant and positive effect on solving reality-based tasks while taking irrelevant notes worsens performance.

To conclude by answering the RQ, *Is there a connection between correctly solving approached reality-based tasks and the qualitatively assessed application of (a) highlighting, (b) writing, (c) evaluating the relevance of information, and (d) elaborating information?* We can say: Yes, the results of the GLMMs indicate a clear connection between the quality of specific note-taking strategies—especially writing out all required information and elaborated notes—and the correct solution for reality-based tasks.

6 Discussion

The findings highlight the importance of the quality of metacognitive strategies in solving complex, reality-based mathematical tasks, particularly in writing information and elaboration. The scientific significance of this study lies in its contribution to understanding how these strategies can be optimized for education contexts.

6.1 Implications of the Findings

Our study employed an experimental design with 1064 student responses. In this way, we contribute to the understanding of learning strategies and their connection to individual competencies such as language and mathematics proficiency. Working with a long dataset at the item level enables targeted analyses and interpretations at the item level.

Regarding RQ1 (*To what extent are the strategies of (a) highlighting, (b) writing, (c) evaluating the relevance of information, and (d) elaborating information applied with varying quality in students' notes during the solution process of reality-based tasks at different language levels?*), it is noticeable that *the students use far fewer organizing strategies, such as highlighting or writing out information, than they do other more elaborate forms of note-taking*. In addition to mathematics and language proficiency, text comprehension, writing out all required information and taking elaborate notes are possible predictors for the correct solution of reality-based tasks.

Regarding RQ2 (*Which personal and task-related language characteristics are associated with creating notes and their quality to derive the solution in reality-based tasks?)* is the finding that better *language proficiency favors the elaboration of relevant information and mathematics proficiency correlates with a lower number of irrelevant notes,* particularly relevant. The ability to distinguish relevant from irrelevant information is linked to proficiency in language and mathematics. We recommend that teachers develop strategies focused on extracting relevant information, particularly for students with weaker language or math skills. Intervention programs could aim to teach students techniques for filtering and structuring information to enhance their learning experience performance.

An important scientific contribution of our study is that it *validates self-reporting of learning strategies,* as people who reported strategic approaches actually implemented them. This strengthens confidence in student self-reports in educational research and suggests that teaching and promoting metacognition, that is, awareness of one's own learning strategies, can benefit learning behavior. Teachers could help students reflect on their learning processes to develop more efficient strategies.

As already confirmed by Slotte et al. (2001) and Murtafi'ah et al. (2023), we found that *girls tend to take more supportive notes* than boys, but that gender still does not correlate with the correct solution. We still recommend gender-sensitive teaching methods to promote more note-taking and information structuring among boys to stress the importance of these strategies for the learning process and, ultimately, close the gaps between the genders.

The results of RQ3 (*Is there a connection between correctly solving approached reality-based tasks and the qualitatively assessed application of (a) highlighting, (b) writing, (c) evaluating the relevance of information, and (d) elaborating information?*) show once again that *learning support* should not only focus on the subject but also on language. The situation is similar for task-related text comprehension—it strongly predicts the correct solution, even with the same general language proficiency. In their study, Leiss et al. (2019) show a strong relationship between the situation model and the solution rate. Since text comprehension is roughly comparable to the situation model, we can include this large predictor in our analysis and say that notes become significant despite explaining the situation model.

Although there are strategies that work *across all tasks, the tasks constructed in parallel show very different solution rates and the use of different types of note-taking*. The notes also show very different correlations to the solution depending on the task context and language level. Regarding the explanation of personal characteristics, in *Task YT*, which had a very low solution rate, we could not prove a correlation between mathematical and language proficiency with the correct solution, which the extant literature defines as a strong predictor (Leiss et al. 2010; Reinhold et al. 2020; Strohmaier et al. 2021). Perhaps the task could be difficult to the extent that even known results do not work here.

Moreover, some predictors have significantly different effects across all tasks despite the strongly fluctuating solution rate, such as writing, avoiding irrelevant notes, and creating notes that indicate deeper cognitive processing, such as resorting the information into an individual order that may be suitable for mathematical calculation (see Table 4, Model 4).

Writing all required information is beneficial for achieving the correct solution (Cox 1999; Samuelstuen and Bråten 2007). However, this study does not clarify whether this is merely a way of relieving the load on working memory (Kiewra 1989; Leutner et al. 2009; Mikk 2008; Sweller 1989) or to achieve a deeper level of processing of the information (Dündar 2015; Graham et al. 2020; Kobayashi 2005; Taylor and McDonald 2007).

Table 4 GLMM with "correct solution" as the dependent variable

	Model 1 Personal Characteristics			Model 2 Task-related predictors			Model 3 Note-Taking Strategies			Model 4 Interaction of Strategies		
Predictors	*Estimate*	*Std. Error*	*p*	*Estimate*	*Std. Error*	*p*	*Estimate*	*Std. Error*	*p*	*Estimate*	*Std. Error*	*p*
Intercept	–1.76***	0.40	**<0.001**	–1.77	1.02	0.082	–1.91	1.09	0.080	–1.97	1.14	0.084
General mathematics proficiency	0.51***	0.11	**<0.001**	0.41***	0.11	**<0.001**	0.41***	0.11	**<0.001**	0.42***	0.12	**<0.001**
General language proficiency	0.80***	0.15	**<0.001**	0.58***	0.14	**<0.001**	0.57***	0.15	**<0.001**	0.60***	0.15	**<0.001**
Cultural resources	0.12	0.10	0.255	0.09	0.10	0.362	0.07	0.10	0.509	0.05	0.11	0.645
Task language level = 2	–	–	–	0.07	0.94	0.940	1.01	0.970	0.05	0.959	0.07	0.94
Task language level = 3	–	–	–	0.01	0.94	0.995	1.01	0.973	0.04	0.969	0.01	0.94
Task-related text comprehension	–	–	–	0.60***	0.11	**<0.001**	0.58***	0.11	**<0.001**	0.57***	0.11	**<0.001**
Highlighting all required information	–	–	–	–	–	–	0.09	0.08	0.301	0.08	0.09	0.333
Writing out all required information	–	–	–	–	–	–	0.23**	0.08	**0.007**	0.22*	0.09	**0.010**
Taking irrelevant notes	–	–	–	–	–	–	–0.35*	0.16	**0.030**	–0.35*	0.16	**0.032**
Taking elaborated notes	–	–	–	–	–	–	0.37***	0.10	**<0.001**	0.32**	0.10	**0.002**
Interaction: comprehension × elaborated notes	–	–	–	–	–	–	–	–	–	0.24*	0.10	**0.019**
Random Effects												
σ^2	3.29			3.29			3.29			3.29		
τ_{00}	0.66 student			0.50 student			0.52 student			0.53 student		
	1.24 task			1.23 task			1.43 task			1.50 task		
N	9 task			9 task			9 task			9 task		
	395 student			395 student			395 student			395 student		
Marginal R^2/Conditional R^2	0.210/0.500			0.244/0.505			0.287/0.553			0.288/0.560		
Deviance	977.766			944.930			913.679			908.096		
AIC	989.766			960.930			937.679			936.096		

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

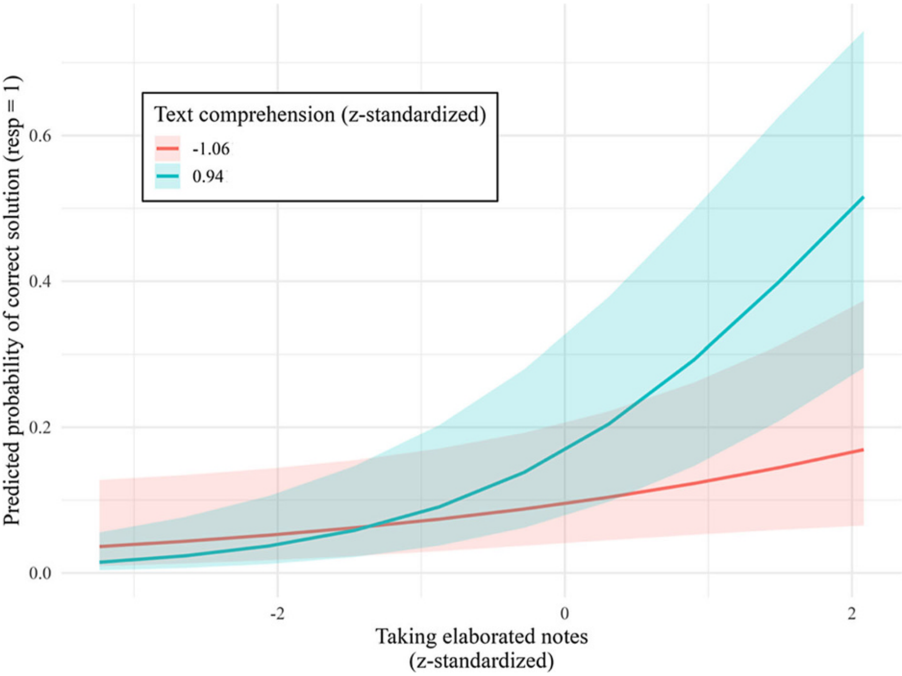


Fig. 6 Interaction Effect between *Text comprehension* and *Taking elaborated notes*

Highlighting all required information does not appear as a significant predictor (see Fig. 6, Model 3), although the recognition of important information is an important step for further work with a text (Azevedo et al. 2013; Bråten and Samuelstuen 2007; Leutner et al. 2007; Schmitz and Dannecker 2023). This non-significant result may also be owing to the present sample's relatively small and unstable highlighting effect. A post-hoc power analysis using 1000 simulations based on the fitted model revealed a statistical power of only 0.8% (95%, confidence interval: 0.35–1.57%) to detect the observed effect of highlighting. This lack of significance, combined with the extremely low statistical power to detect the effect of highlighting (power = 0.8%), suggests that students may not have used the strategy effectively or purposefully in this context. Rather than indicating a general inefficacy of highlighting, the result may reflect the need for students to apply more goal-oriented and adaptive strategies, depending on the task's demands (Bråten and Samuelstuen 2007; Hadwin et al. 2001). Both writing and highlighting may organize the reading and solving process (Weinstein and Mayer 1986).

Interestingly, at *the most difficult language level (LL3), organization strategies become significant compared with elaboration* (see Table 4, Model 4). As the linguistic demand increases, it becomes more difficult to correctly assess the relevance of the information, making the organization more important. The general importance of organizing information (Leutner et al. 2009) is also reflected in correctly classifying relevant and irrelevant information to solve the task (see Table 4, Model 4). A more concentrated effort could be devoted to instructing students on accurately categorizing information according to relevance. As anticipated, elaboration emerged as the most significant predictor, since elaborative notes reflect a more profound comprehension of the topic task (Rogiers et al. 2020; Rovers et al. 2019; Schmitz and Karstens 2022; Weinstein and Mayer 1986).

The *interaction effect of the GLMM* (see Table 4, Model 4) shows that the probability of a correct solution increases with increasing elaborated notes for people with high and low text comprehension. However, the slope of the group with high text comprehension is steeper, indicating that this group benefits more from elaborated notes. One explanation for this effect could be that elaborated notes, which require a higher processing depth, also demand a higher level of text comprehension. Individuals with a high level of text comprehension can better assimilate the noted information into the solution process, significantly increasing their chances of arriving at a correct solution. These results highlight the importance of fostering metacognitive awareness and teaching students to use strategies purposefully rather than routinely. Incorporating such strategy training into teacher education and classroom practices can help students better understand texts and enhance their problem-solving skills in complex, reality-based tasks. Ultimately, by bridging insights from educational research with instructional practice, this study provides a strong empirical foundation for supporting strategic learning behavior. Helping students learn how to learn—especially through targeted and thus high-quality note-taking and information-processing strategies—may be key to empowering them in diverse mathematical tasks within real contexts and beyond.

6.2 Limitations and Future Research Directions

This study offers valuable insights into how supportive note-taking strategies impact the resolution of real-life mathematical tasks. Nevertheless, it's important to acknowledge certain limitations to contextualize the results and their generalizability. A significant limitation is that the findings stem from a specific convenience sample, potentially restricting their applicability to other groups or educational settings. Additionally, the limited variety of task types and linguistic features narrows the scope of the findings. We encourage researchers to investigate a wider array of tasks and linguistic complexities to enhance understanding of how different task characteristics interact with supportive note-taking strategies and influence solving effectiveness. This approach will yield a more thorough understanding of the relationship between task difficulty and cognitive processes.

Moreover, the relevance of note-taking diminishes significantly when it comes to text comprehension, which involves a deeper engagement with the relevant information. Students who have addressed nine content-related questions in advance might assume that they only need organizational strategies like highlighting or writing less in a situation where no prior questions are presented. However, Table 2 indicates positive correlations with text comprehension, revealing that higher comprehension levels are associated with more extensive and high-quality notes. Nevertheless, the comprehension questions act as an intervention, suggesting that the findings do not reflect the "pure" solution process of reality-based tasks. Consequently, only a few item properties and characteristics were analyzable at the item level, which was inadequate for establishing clearly interpretable correlations and regressions. Still, we observed significant differences between the tasks, despite their parallel construction, and potential factors contributing to this have been discussed in Sects. 4 and 5, while recognizing the need for clearer conclusions. Despite this intervention and the pronounced differences between the similarly constructed tasks, note-taking still creates a notable impact, as evidenced by the model comparison shows.

Another limitation is the calculation of predictors across all age groups. Some studies (e.g., Faber et al. 2000) argue that taking notes is more cognitively demanding than solving the task or learning the content. Likewise, the additional cognitive burden of drawing the sketch can hinder the solution and comprehension process as Krawitz and Schukajlow (2020) show, since the participants focused on mathematically irrelevant aspects. They show that sketches have no or even negative effects on the solution processes (Krawitz and Schukajlow 2020). This study refers in particular to De Bock et al. (2002), who surprisingly demonstrated a negative relation of drawing on the performance of solving nonlinear geometry problems. They argue that this finding is possibly attributable to misconceptions about the basic mathematical concepts (vom Hofe and Blum 2016), which a sketch could reinforce. However, as previously indicated, the effects of solution strategies depend on the mathematical topic (e.g., geometry or algebra; Schukajlow et al. 2015).

Additionally, De Corte and Verschaffel (1987) also show that students think significantly more than they take notes. Our study does not cover all elaborative cognitive thoughts, as specific learners do not need notes and can mentally elaborate on the situation.

We only considered "all required information" codes for writing and highlighting to exclude random underlining and notes. However, we also included the quality of the notes so that these two codes primarily indicate that students who can determine the required information in their notes performed better on the tasks. Furthermore, the increasing integration of digital tools and environments in education presents an opportunity to examine and explain the usage of note-taking strategies. Investigating how technology, such as digital note-taking applications or collaborative platforms, shapes the use and effectiveness of these strategies could enhance the practical relevance of our findings and support the development of innovative instructional approaches. As such, reality-based tasks, which aim to mirror real-world problems, present higher cognitive demands but offer opportunities for students to apply mathematics in authentic contexts. These tasks, however, tend to have lower solution rates than traditional math problems. Our findings highlight the importance of the quality and depth of processing of different types of note-taking—such as relevant, irrelevant, and elaborate notes—in successfully solving reality-based tasks. Effective note-taking helps students better comprehend and tackle real-world problems through mathematics. To increase the practical relevance of these findings, future research should explore targeted interventions that promote effective note-taking strategies. Teaching methods focused on identifying relevant information and elaboration, integrated into teacher training and classroom materials, could improve the solution process and strategic learning outcomes.

Furthermore, the connection between research and practice ensures that educational theory is translated into meaningful classroom practice. However, owing to the experimental focus of this study, we must establish the effectiveness of specific interventions before advocating for widespread implementation. Future research should investigate how task-specific strategies can be adapted to task-specific difficulties to add sustainable value to teaching and learning.

In conclusion, this study highlights the critical role of note-taking strategies and the quality of notes in empowering students to face real-world challenges. To further support this conclusion, future studies should consider using task-specific strategies tailored to address specific task difficulties. This approach could enhance our understanding of how note-taking supports modern education and lifelong learning.

Supplementary Information The online version of this article (https://doi.org/10.1007/s13138-025-00263-x) contains supplementary material, which is available to authorized users.

Funding The VAMPS project is a funded by German Research Foundation (Variation Tasks Mathematics Physics Language, project No. 417017613, project leaders: Dominik Leiss, Knut Schwippert, Timo Ehmke, Lena Heine, Dietmar Höttecke).

Funding Open Access funding enabled and organized by Projekt DEAL.

Conflict of interest L.-M. Wienecke, D. Leiss and T. Ehmke declare that they have no competing interests.

References

Adams, T. L. (2003). Reading mathematics: more than words can say. *Reading Teacher*, *56*(8), 786–795.

Aebli, H., Staub, F. C., & Ruthemann, U. (1991). Textrechnungen im Mathematikunterricht. *Mathematik Lehren*, *44*, 12–17.

Afflerbach, P., Hurt, M., & Cho, B.-Y. (2020). Reading comprehension strategy instruction. In D. L. Dinsmore, L. K. Fryer & M. M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 98–118). Routledge. https://doi.org/10.4324/9780429423635-7.

Anderson, J. (1981). *Analysing the readability of English and non-English texts in the classroom with LIX*. Paper Presented at the Annual Meeting of the Australian Reading Association, Darwin.

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427–449). New York: Springer. https://doi.org/10.1007/978-1-4419-5546-3_28.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.i01.

Beaudine, G. (2022). Mathematical reading: Investigating the reading xomprehension strategies implemented by middle school students. *International Journal of Science and Mathematics Education*, *20*(Suppl. 1), 187–213. https://doi.org/10.1007/s10763-022-10287-1.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
Berkowitz, M., & Stern, E. (2018). Which cognitive abilities make the difference? Predicting academic achievements in advanced STEM studies. *Journal of Intelligence*, *6*(4), 4. https://doi.org/10.3390/jintelligence6040048.
Bloom, B. S. (1956). Taxonomy of educational objectives: the classification of educational goals. In *Cognitive domain*.
Blum, W., & Leiss, D. (2007). How do students and teachers deal with modelling problems. Mathematical modelling (ICTMA 12): Education, engineering and economics, 222–231.
Blum, W., Galbraith, P. L., Henn, H.-W., & Niss, M. (Eds.). (2007). *Modelling and applications in mathematics education: the 14th ICMI Study*. Vol. 10. Springer. https://doi.org/10.1007/978-0-387-29822-1.
Boaler, J. (2001). Mathematical modelling and new theories of learning. *Teaching Mathematics and Its Applications*, *20*(3), 121–128. https://doi.org/10.1093/teamat/20.3.121.
Board, U. S. N. A. G. (2010). *Mathematics framework for the 2011 National Assessment of Educational Progress*. National Assessment Governing Board, United States Department of Education.
Boeck, P. D., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v039.i12.
Bohn-Gettler, C. M., & Kendeou, P. (2014). The interplay of reader goals, working memory, and text structure during reading. *Contemporary Educational Psychology*, *39*(3), 206–219. https://doi.org/10.1016/j.cedpsych.2014.05.003.
Borromeo Ferri, R. (2007). Individual modelling routes of pupils—analysis of modelling problems in mathematical lessons from a cognitive perspective. In C. Haines (Ed.), *Mathematical modelling (ICTMA 12): education, engineering and economics* (pp. 260–270). Horwood.
Bråten, I., & Samuelstuen, M. S. (2007). Measuring strategic processing: comparing task-specific self-reports to traces. *Metacognition and Learning*, *2*(1), 1–20. https://doi.org/10.1007/s11409-007-9004-y.
Bretzing, B. H., & Kulhavy, R. W. (1979). Notetaking and depth of processing. *Contemporary Educational Psychology*, *4*(2), 145–153. https://doi.org/10.1016/0361-476X(79)90069-9.
Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644.
Buschmeier, G. (2017). *Denken und Rechnen (Thinking and Calculating)* (2nd edn.). Westermann. (Grundschule, Bremen, Hamburg, Hessen, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Schleswig-Holstein), Druck A
Chang, W.-C., & Ku, Y.-M. (2015). The effects of note-taking skills instruction on elementary students' reading. *Journal of Educational Research*, *108*(4), 278–291. https://doi.org/10.1080/00220671.2014.886175.
Clarkson, P. (1991). Language comprehension errors: a further investigation. *Mathematics Education Research Journal*, *3*(2), 24–33. https://doi.org/10.1007/BF03217225.
Corte, E. D., & Verschaffel, L. (1987). The effect of semantic structure on first graders' strategies for solving addition and subtraction word problems. *Journal for Research in Mathematics Education*, *18*(5), 363–381. https://doi.org/10.5951/jresematheduc.18.5.0363.
Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, *9*(4), 343–363. https://doi.org/10.1016/S0959-4752(98)00051-6.
Cumming, G. (2013). The New Statistics: Why and How. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966. Original work published 2014.
Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, *20*(4), 405–438. https://doi.org/10.1016/0010-0285(88)90011-4.
Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: a review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, *6*, 348. https://doi.org/10.3389/fpsyg.2015.00348.
De Bock, D., Verschaffel, L., & Janssens, D. (2002). The effects of different problem presentations and formulations on the illusion of linearity in secondary school students. *Mathematical Thinking and Learning*, *4*(1), 65–89. https://doi.org/10.1207/S15327833MTL0401_3.
De Bock, D., Verschaffel, L., Janssens, D., Van Dooren, W., & Claes, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative

evidence from a study on modelling non-linear geometry problems. *Learning and Instruction*, *13*(4), 441–463. https://doi.org/10.1016/S0959-4752(02)00040-3.

Degrande, T., Van Hoof, J., Verschaffel, L., & Van Dooren, W. (2018). Open word problems: taking the additive or the multiplicative road? *ZDM*, *50*(1–2), 91–102. https://doi.org/10.1007/s11858-017-0900-6.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Ding, H., & Homer, M. (2020). Interpreting mathematics performance in PISA: taking account of reading performance. *International Journal of Educational Research*, *102*, 101566. https://doi.org/10.1016/j.ijer.2020.101566.

Dinsmore, D. L., Fryer, L. K., & Parkinson, M. M. (Eds.). (2020). *Handbook of strategies and strategic processing*. Routledge.

Dröse, J., & Prediger, S. (2018). Strategien für Textaufgaben – Fördern mit Info-Netzen und Formulierungsvariationen. *Mathematik Lehren*, *206*, 8–12.

Dumas, D. (2020). Strategic processing within and across domains of learning. In D. L. Dinsmore, L. K. Fryer & M. M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 11–28). Routledge. https://doi.org/10.4324/9780429423635-2.

Dündar, S. (2015). Does writing have any effect on mathematics success? *Journal of Education and Training Studies*, *4*(1), 1–10. https://doi.org/10.11114/jets.v4i1.989.

El Masri, Y. H., Ferrara, S., Foltz, P. W., & Baird, J.-A. (2017). Predicting item difficulty of science national curriculum tests: The case of key stage 2 assessments. *Curriculum Journal*, *28*(1), 59–82. https://doi.org/10.1080/09585176.2016.1232201.

Embretson, S. E., & Daniel, R. C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. *Psychology Science*, *50*(3), 328.

English, L., & Lesh, R. (2003). Ends-in-view problems. In *Beyond constructivism* (pp. 297–316).

Faber, J. E., Morris, J. D., & Lieberman, M. G. (2000). The effect of note taking on ninth grade students' comprehension. *Reading Psychology*, *21*(3), 257–270. https://doi.org/10.1080/02702710050144377.

Ferrara, S., Svetina, D., Skucha, S., & Davidson, A. H. (2011). Test development with performance standards and achievement growth in mind. *Educational Measurement: Issues and Practice*, *30*(4), 3–15. https://doi.org/10.1111/j.1745-3992.2011.00218.x.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, *98*(1), 1. https://doi.org/10.1037/0022-0663.98.1.29.

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Martin, B. N. (2018). Text comprehension and oral language as predictors of word-problem dolving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading*, *22*(2), 152–166. https://doi.org/10.1080/10888438.2017.1398259.

Galbraith, P., & Stillman, G. (2006). A framework for identifying student blockages during transitions in the modelling process. *ZDM*, *38*(2), 143–162. https://doi.org/10.1007/BF02655886.

Gijsbers, D., de Putter-Smits, L., & Pepin, B. (2020). Changing students' beliefs about the relevance of mathematics in an advanced secondary mathematics class. *International Journal of Mathematical Education in Science and Technology*, *51*(1), 87–102. https://doi.org/10.1080/0020739X.2019.1682698.

Goos, M., Carreira, S., & Namukasa, I. K. (2023). Mathematics and interdisciplinary STEM education: recent developments and future directions. *ZDM*, *55*(7), 1199–1217. https://doi.org/10.1007/s11858-023-01533-z.

Graham, S., Kiuhara, S. A., & MacKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: a meta-analysis. *Review of Educational Research*, *90*(2), 179–226. https://doi.org/10.3102/0034654320914744.

Gravemeijer, K., Stephan, M., Julie, C., Lin, F.-L., & Ohtani, M. (2017). What mathematics education may prepare students for the society of the future? *International Journal of Science and Mathematics Education*, *15*(Suppl. 1), 105–123. https://doi.org/10.1007/s10763-017-9814-6.

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504.

Grotjahn, R., & Drackert, A. (2020). *The electronic C-Test bibliography: Version October 2020* (p. 66).

Grotjahn, R., Klein-Barley, C., & Raatz, U. (2002). C-test: an overview. In J. A. Coleman, R. Grotjahn & U. Raatz (Eds.), *University language testing and the C-Test* (pp. 93–114). AKS.

Haag, N., Heppt, B., Roppelt, A., & Stanat, P. (2015). Linguistic simplification of mathematics items: effects for language minority students in Germany. *European Journal of Psychology of Education*, *30*(2), 145–167. https://doi.org/10.1007/s10212-014-0233-6.

Hackemann, T., Heine, L., & Höttecke, D. (2022). Challenging to read, easy to comprehend? Effects of linguistic demands on secondary students' text comprehension in physics. *International Journal of Science and Mathematics Education*, *20*(Suppl. 1), 43–68. https://doi.org/10.1007/s10763-022-10306-1.

Hadwin, A.F., Winne, P.H., Stockley, D.B., Nesbit, J.C., & Woszczyna, C. (2001). Context moderates students' self-reports about how they study. *Journal of Educational Psychology*, *93*(3), 477–487. https://doi.org/10.1037/0022-0663.93.3.477.

Harsch, C., & Hartig, J. (2016). Comparing C-tests and yes/no vocabulary size tests as predictors of receptive language skills. *Language Testing*, *33*(4), 555–575. https://doi.org/10.1177/0265532215594642.

Härtig, H., Heitmann, P., & Retelsdorf, J. (2015). Analyses of the tasks for evaluating the educational standards in physics—Differentiation between written language proficiency and content knowledge. *Zeitschrift für Erziehungswissenschaft*, *18*(4), 763–779. https://doi.org/10.1007/s11618-015-0646-2.

Härtig, H., Fraser, N., Bernholt, S., & Retelsdorf, J. (2019). Kann man Sachtexte vereinfachen? – Ergebnisse einer Generalisierungsstudie zum Textverständnis (Can factual texts be simplified?—Results of a generalization study on text comprehension). *Zeitschrift für Didaktik der Naturwissenschaften*, *25*(1), 273–287. https://doi.org/10.1007/s40573-019-00105-7.

Heine, L., Domenech, M., Otto, L., Neumann, A., Krelle, M., Leiss, D., Höttecke, D., Ehmke, T., & Schwippert, K. (2018). Modellierung sprachlicher Anforderungen in Testaufgaben verschiedener Unterrichtsfächer: Theoretische und empirische Grundlagen (Modeling linguistic requirements in test tasks for various school subjects: Theoretical and empirical foundations). *Zeitschrift für Angewandte Linguistik*, *2018*(69), 69–96. https://doi.org/10.1515/zfal-2018-0017.

vom Hofe, R., & Blum, W. (2016). 'Grundvorstellungen' as a category of subject-matter didactics. *Journal für Mathematik-Didaktik*, *37*(Suppl. 1), 225–254. https://doi.org/10.1007/s13138-016-0107-3.

Johnson, E., & Monroe, B. (2004). Simplified language as an accommodation on math tests. *Assessment for Effective Intervention*, *29*(3), 35–45. https://doi.org/10.1177/073724770402900303.

Julie, C. (2007). Learners' context preferences and mathematical literacy. In *Mathematical modelling (ICTMA12): Education, engineering and economics* (pp. 195–220). https://doi.org/10.1533/9780857099419.4.195.

Kaiser, G. (2017). The teaching and learning of mathematical modelling. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 267–291). National Council of Teachers of Mathematics.

Kaiser, G., Blum, W., Stillman, G., & Borromeo Ferri, R. (Eds.). (2011). *Trends in teaching and learning of mathematical modelling: ICTMA 14. International Conference on the Teaching of Mathematical Modelling and Applications*. Dordrecht, New York: Springer.

Karstens, F., & Schmitz, A. (2023). *Lesestrategien zur Unterstützung des Textverstehens im sprachbewussten Biologieunterricht: Empirische Einblicke aus der Sicht von Lernenden*. Publisher: Verein Leseforum. Reading strategies to support text comprehension in language-aware biology lessons: Empirical insights from the perspective of learners

Kiewra, K.A. (1989). A review of note-taking: the encoding-storage paradigm and beyond. *Educational Psychology Review*, *1*(2), 147–172. https://doi.org/10.1007/BF01326640.

Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363.

Kintsch, W., & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, *92*(1), 109–129. https://doi.org/10.1037/0033-295X.92.1.109.

Klotz, E., Ehmke, T., & Leiss, D. (2025). Text comprehension as a mediator in solving mathematical reality-based tasks: the impact of linguistic complexity, cognitive factors, and social background. *European Journal of Educational Research*, *14*(1), 23–39. https://doi.org/10.12973/eu-jer.14.1.23.

Knabbe, A., Leiss, D., & Ehmke, T. (2025). Reality-based tasks with complex-situations: identifying sociodemographic and cognitive factors for solution. *International Journal of Science and Mathematics Education*, *23*(1), 97–120. https://doi.org/10.1007/s10763-024-10463-5.

Kobayashi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology*, *30*(2), 242–262. https://doi.org/10.1016/j.cedpsych.2004.10.001.

Krawitz, J., & Schukajlow, S. (2020). When can making a drawing hinder problem solving? Effect of the drawing strategy on linear overgeneralizations and problem solving. *Frontiers in Psychology*, *11*, 506. https://doi.org/10.3389/fpsyg.2020.00506.

Krawitz, J., Schukajlow, S., Chang, Y.-P., & Yang, K.-L. (2017). *Reading comprehension, enjoyment, and performance in solving modelling problems: How important is a deeper situation model?*

Krulik, S., & Rudnick, J. A. (1996). *The new sourcebook for teaching reasoning and problem solving in junior and senior high school. (No Title)*

Lahtinen, V., Lonka, K., & Lindblom-Ylänne, S. (1997). Spontaneous study strategies and the quality of knowledge construction. *British Journal of Educational Psychology*, *67*(1), 13–24. https://doi.org/10.1111/j.2044-8279.1997.tb01223.x.

Leiss, D., & Tropper, N. (2014). *Umgang mit Heterogenität im Mathematikunterricht: Adaptives Lehrerhandeln beim Modellieren (Dealing with heterogeneity in mathematics teaching: Adaptive teacher behavior in modeling)*. Springer Spektrum.

Leiss, D., Schukajlow, S., Blum, W., Messner, R., & Pekrun, R. (2010). The role of the situation model in mathematical modelling—task analyses, student competencies, and teacher interventions. *Journal Für Mathematik-Didaktik*, *31*(1), 119–141. https://doi.org/10.1007/s13138-010-0006-y.

Leiss, D., Plath, J., & Schwippert, K. (2019). Language and mathematics—key factors influencing the comprehension process in reality-based tasks. *Mathematical Thinking and Learning*, *21*(2), 131–153. https://doi.org/10.1080/10986065.2019.1570835.

Leiss, D., Ehmke, T., & Heine, L. (2024). Reality-based tasks for competency-based education: the need for an integrated analysis of subject-specific, linguistic, and contextual task features. *Learning and Individual Differences*, *114*, 102518. https://doi.org/10.1016/j.lindif.2024.102518.

Leopold, C., & Leutner, D. (2002). Der Einsatz von Lernstrategien in einer konkreten Lernsituation bei Schülern unterschiedlicher Jahrgangsstufen (The use of learning strategies in a specific learning situation among students of different grade levels). *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen (Quality of school education: School and non-school conditions for mathematical, scientific, and interdisciplinary skill)*, 240–258.

Lester, F. K., Garofalo, J., & Kroll, D. L. (1989). *The Role of metacognition in mathematical problem solving: A Study of two grade seven classes*. Educational Resources Information Center.

Leutner, D., Leopold, C., & Den Elzen-Rump, V. (2007). Self-regulated learning with a text-highlighting strategy. *Zeitschrift für Psychologie/Journal of Psychology*, *215*(3), 174–182. https://doi.org/10.1027/0044-3409.215.3.174.

Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: effects of drawing and mentally imagining text content. *Computers in Human Behavior*, *25*(2), 284–289. https://doi.org/10.1016/j.chb.2008.12.010.

Lewalter, D., Diedrich, J., Goldhammer, F., Köller, O., & Reiss, K. (Eds.). (2023). *PISA 2022: Analyse der Bildungsergebnisse in Deutschland*. Waxmann. https://doi.org/10.31244/9783830998488.

Lin, K.-C., & Chen, Y.-J. (2015). Detecting misspecification in the random-effects structure of cumulative logit models. *Computational Statistics & Data Analysis*, *92*, 126–133. https://doi.org/10.1016/j.csda.2015.07.002.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological Bulletin*, *136*(6), 6. https://doi.org/10.1037/a0021276.

Littlefield, J., & Rieser, J. J. (1993). Semantic features of similarity and children's strategies for identifying relevant information in mathematical story problems. *Cognition and Instruction*, *11*(2), 133–188. https://doi.org/10.1207/s1532690xci1102_2.

Lonka, K., Lindblom-Ylänne, S., & Maury, S. (1994). The effect of study strategies on learning from text. *Learning and Instruction*, *4*(3), 253–271. https://doi.org/10.1016/0959-4752(94)90026-4.

Lu, X., & Huang, J. (2021). Mathematical modelling in China: How it is described and required in mathematical curricula and what is the status of students' performance on modelling tasks. In B. Xu, Y. Zhu & X. Lu (Eds.), *Beyond Shanghai and Pisa: Cognitive and non-cognitive competencies of Chinese students in mathematics* (pp. 209–233). Springer. https://doi.org/10.1007/978-3-030-68157-9_12.

Lubienski, S. T. (2000). Problem solving as a means toward mathematics for all: an exploratory look through a class lens. *Journal for Research in Mathematics Education*, *31*(4), 4. https://doi.org/10.2307/749653.

Maaß, K. (2006). What are modelling competencies? *ZDM*, *38*(2), 2. https://doi.org/10.1007/BF02655885.

Mang, J., Ustjanzew, N., Schiepe-Tiska, A., Prenzel, M., Sälzer, C., Müller, K., & Rodríguez, G. E. (2018). *PISA 2012 Skalenhandbuch: Dokumentation der Erhebungsinstrumente (PISA 2012 Scale Manual: Documentation of the survey instruments)*. Waxmann.

Manzi, A., Martinez, S., & Durmysheva, Y. (2017). Cognitive correlates of lecture note taking: handwriting speed and attention. *North American Journal of Psychology*, *19*(1)., .

Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives* (2nd edn.). Corwin Press.

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Bem-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29–54). Lawrens Erlbaum.
McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, *51*, 297–384. https://doi.org/10.1016/S0079-7421(09)51009-2.
Mevarech, Z. R., Terkieltaub, S., Vinberger, T., & Nevet, V. (2010). The effects of meta-cognitive instruction on third and sixth graders solving word problems. *ZDM*, *42*(2), 195–203. https://doi.org/10.1007/s11858-010-0244-y.
Mikk, J. (2008). Sentence length for revealing the cognitive load reversal effect in text comprehension. *Educational Studies*, *34*(2), 2. https://doi.org/10.1080/03055690701811164.
Montague, M., Krawec, J., Enders, C., & Dietz, S. (2014). The effects of cognitive strategy instruction on math problem solving of middle-school students of varying ability. *Journal of Educational Psychology*, *106*(2), 469–481. https://doi.org/10.1037/a0035176.
Mullis, I. V. S., Martin, M. O., Fishbein, B., Foy, P., & Moncaleano, S. (2020). *Findings from the TIMSS 2019 problem solving and inquiry tasks*. Boston College and IAE: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and IAE.
Mullis, I. V. S., Martin, M. O., & von Davier, M. (Eds.). (2021). *TIMSS 2023 assessment frameworks*. International Association for the Evaluation of Educational Achievement. 1017 BT
Murtafi'ah, M., Muflihah, M., & Zen, E. L. (2023). Note-taking: the gender difference on students' perception. *Scope*, *7*(2), 226. https://doi.org/10.30998/scope.v7i2.15361.
Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intraclass correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 134. https://doi.org/10.1098/rsif.2017.0213.
Nanda, A., Mohapatra, D. B. B., Mahapatra, A. P. K., Mahapatra, A. P. K., & Mahapatra, A. P. K. (2021). Multiple comparison test by Tukey's honestly significant difference (HSD): do the confident level control type I error. *International Journal of Statistics and Applied Mathematics*, *6*(1), 59–65. https://doi.org/10.22271/maths.2021.v6.i1a.636.
National Council of Teachers of Mathematics (2000). Principles and standards for school mathematics. National Council of Teachers of Mathematics. https://www.nctm.org/Standards-and-Positions/Principles-and-Standards
Neri, C. N., & Retelsdorf, J. (2022). The role of linguistic features in science and math comprehension and performance: a systematic review and desiderata for future research. *Educational Research Review*, *36*, 100460. https://doi.org/10.1016/j.edurev.2022.100460.
Nesher, P., & Teubal, E. (1975). Verbal cues as an interfering factor in verbal problem solving. *Educational Studies in Mathematics*, *6*(1), 1. https://doi.org/10.1007/BF00590023.
Newton, K. (2020). Mathematics strategy interventions. In D. L. Dinsmore, L. K. Fryer & M. M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 159–176). Routledge. https://doi.org/10.4324/9780429423635-10.
Niss, M., & Blum, W. (2020). *The learning and teaching of mathematical modelling* (1st edn.). Routledge. https://doi.org/10.4324/9781315189314.
Niss, M., Blum, W., & Galbraith, P. L. (2007). Introduction. In W. Blum, P. L. Galbraith, H.-W. Henn & M. Niss (Eds.), *Modelling and applications in mathematics education: the 14th ICMI study* 1st edn. Springer.
Nurjanah, A., & Retnowati, E. (2018). Analyzing the extraneous cognitive load of a 7th grader mathematics textbook. *Journal of Physics: Conference Series*, *1097*, 12131. https://doi.org/10.1088/1742-6596/1097/1/012131.
Organization for Economic Co-operation and Development (2013). PISA 2012 results: excellence through equity. In *Giving every student the chance to succeed* Vol. II. OECD Publishing. https://doi.org/10.1787/9789264201132-en.
Organization for Economic Co-operation and Development (2019). *PISA 2018 results. What students know and can do: Vol. PISA*. Vol. I. OECD: OECD Publishing. https://doi.org/10.1787/5f07c754-en.
Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 3. https://doi.org/10.1016/j.learninstruc.2008.04.003.
Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1.
Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: a meta-analysis. *Psychological Bulletin*, *146*(7), 595–634. https://doi.org/10.1037/bul0000231.

Piel, S., & Schuchart, C. (2014). Social origin and success in answering mathematical word problems: the role of everyday knowledge. *International Journal of Educational Research*, *66*, 22–34. https://doi.org/10.1016/j.ijer.2014.02.003.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. https://doi.org/10.1177/0013164493053003024.

Plath, J., & Leiss, D. (2018). The impact of linguistic complexity on the solution of mathematical modelling tasks. *ZDM*, *50*(1–2), 159–171. https://doi.org/10.1007/s11858-017-0897-x.

Pollak, H. O. (1977). The interaction between mathematics and other school subjects (including integrated courses). In *Proceedings of the Third International Congress on Mathematical Education, Zentralblatt für Didaktik der Mathematik* (pp. 255–264).

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM*, *52*(1), 33–44. https://doi.org/10.1007/s11858-019-01118-9.

Prediger, S. (2017). Auf sprachliche Heterogenität im Mathematikunterricht vorbereiten (Preparing for linguistic diversity in mathematics lessons). In J. Leuders, T. Leuders, S. Prediger & S. Ruwisch (Eds.), *Mit Heterogenität im Mathematikunterricht umgehen lernen (Learning to deal with heterogeneity in mathematics teaching)* (pp. 29–40). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-16903-9_3.

Prediger, S., & Krägeloh, N. (2015). Low achieving eighth graders learn to crack word problems: a design research project for aligning a strategic scaffolding tool to students' mental processes. *ZDM*, *47*(6), 947–962. https://doi.org/10.1007/s11858-015-0702-7.

Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E., & Benholz, C. (2018). Language proficiency and mathematics achievement: empirical study of language-induced obstacles in a high stakes test, the central exam ZP10. *Journal für Mathematikdidaktik*, *39*(S1), 1–26. https://doi.org/10.1007/s13138-018-0126-3.

Pressley, M., & Afflerbach, P. (2012). *Verbal protocols of reading: the nature of constructively responsive reading*. Routledge.

Reddington, L. A., Peverly, S. T., & Block, C. J. (2015). An examination of some of the cognitive and motivation variables related to gender differences in lecture note-taking. *Reading and Writing*, *28*(8), 1155–1185. https://doi.org/10.1007/s11145-015-9566-z.

Reinhold, F., Hofer, S., Berkowitz, M., Strohmaier, A., Scheuerer, S., Loch, F., Vogel-Heuser, B., & Reiss, K. (2020). The role of spatial, verbal, numerical, and general reasoning abilities in complex word problem solving for young female and male adults. *Mathematics Education Research Journal*, *32*(2), 189–211. https://doi.org/10.1007/s13394-020-00331-0.

Rellensmann, J., Schukajlow, S., & Leopold, C. (2020). Measuring and investigating strategic knowledge about drawing to solve geometry modelling problems. *ZDM*, *52*(1), 97–110. https://doi.org/10.1007/s11858-019-01085-1.

Reusser, K. (1997). Erwerb mathematischer Kompetenzen: Literaturüberblick (Acquisition of mathematical skills: literature review). In F. E. Weinert & A. Helmke (Eds.), *Entwicklung im Grundschulalter (Development in elementary school age)* (pp. 141–156). Beltz, Psychologie-Verl.-Union.

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: a meta-analysis and instructional design principles. *ZDM*, *49*(4), 599–611. https://doi.org/10.1007/s11858-017-0834-z.

Rogiers, A., Merchie, E., & Van Keer, H. (2020). What they say is what they do? Comparing task-specific self-reports, think-aloud protocols, and study traces for measuring secondary school students' text-learning strategies. *European Journal of Psychology of Education*, *35*(2), 315–332. https://doi.org/10.1007/s10212-019-00429-5.

Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2019). Granularity matters: comparing different ways of measuring self-regulated learning. *Metacognition and Learning*, *14*(1), 1–19. https://doi.org/10.1007/s11409-019-09188-6.

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *British Journal of Educational Psychology*, *77*(2), 351–378. https://doi.org/10.1348/000709906X106147.

Schilcher, A., Krauss, S., Kirchhoff, P., Lindl, A., Hilbert, S., Asen-Molz, K., Ehras, C., Elmer, M., Frei, M., Gaier, L., Gastl-Pischetsrieder, M., Gunga, E., Murmann, R., Röhrl, S., Ruck, A.-M., Weich, M., Dittmer, A., Fricke, M., Hofmann, B., & Thim-Mabrey, C. (2021). FALKE: Experiences From

Transdisciplinary Educational Research by Fourteen Disciplines. *Frontiers in Education*, *5*, 579982. https://doi.org/10.3389/feduc.2020.579982.

Schlüter, D., & Besser, M. (2024). Assessing authenticity in modelling test items: deriving a theoretical model. *Frontiers in Education*, *9*, 1343510. https://doi.org/10.3389/feduc.2024.1343510.

Schmidt, S., Ennemoser, M., & Krajewski, K. (2012). *DEMAT 9. Deutscher Mathematiktest für neunte Klassen (German mathematics test for ninth graders)*. Hogrefe.

Schmitz, A., & Dannecker, W. (2023). Strategies for expository and literary texts: students' perspectives on text type-specific cognitive and metacognitive reading strategies. *L1-Educational Studies in Language and Literature. Strategies for Expository and Literary Texts*, *23*, 1–22. https://doi.org/10.21248/l1esll.2023.23.1.407.

Schmitz, A., & Karstens, F. (2022). Lesestrategien zur Unterstützung des Verstehens von Textaufgaben. Vermittlung und Routinen im Mathematikunterricht aus Sicht von Lehrkräften und Lernenden. (Reading strategies to support understanding of text-based tasks. Teaching and routines in mathematics lessons from the perspective of teachers and learners). *Journal für Mathematik-Didaktik*. https://doi.org/10.1007/s13138-021-00188-1.

Schnotz, W., & Dutke, S. (2004). *Kognitionspsychologische Grundlagen der Lesekompetenz: Mehrebenenverarbeitung anhand multipler Informationsquellen (Cognitive psychological foundations of reading competence: multilevel processing using multiple sources of information)*. VS. Verlag für Sozialwissenschaften.

Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics*, *79*(2), 215–237. https://doi.org/10.1007/s10649-011-9341-2.

Schukajlow, S., Krug, A., & Rakoczy, K. (2015). Effects of prompting multiple solutions for modelling problems on students' performance. *Educational Studies in Mathematics*, *89*(3), 3. https://doi.org/10.1007/s10649-015-9608-0.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology. General*, *117*(3), 258258–258275. https://doi.org/10.1037/0096-3445.117.3.258.

Slotte, V., & Lonka, K. (1999). Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology*, *24*(1), 1–20. https://doi.org/10.1006/ceps.1998.0980.

Slotte, V., Lonka, K., & Lindblom-Ylänne, S. (2001). Study-strategy use in learning from text. Does gender make any difference? *Instructional Science*, *29*(3), 255–272. https://doi.org/10.1023/A:1017574300304.

Staub, F. C. (2006). *Notizenmachen: Funktionen, Formen und Werkzeugcharakter von Notizen (Note-taking: Functions, forms, and tools for taking notes)* (pp. 59–71).

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: an analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488. https://doi.org/10.3102/00028312033002455.

Stender, P. (2018). The use of heuristic strategies in modelling activities. ZDM. *Article*, *50*(1), 1. https://doi.org/10.1007/s11858-017-0901-5.

Stern, E. (1992). Die spontane Strategieentdeckung in der Arithmetik (The spontaneous discovery of strategies in arithmetic). In: *Lern- und Denkstrategien. Analyse und Intervention (Learning and thinking strategies. Analysis and intervention)* (pp. 101–123).

Stillman, G., Brown, J., & Galbraith, P. (2010). Identifying challenges within transition phases of mathematical modeling activities at year 9. In R. Lesh, P. L. Galbraith, C. R. Haines & A. Hurford (Eds.), *Modeling students' mathematical modeling competencies* (pp. 385–398). Springer. https://doi.org/10.1007/978-1-4419-0561-1_33.

Stillman, G. A., Kaiser, G., Blum, W., & Brown, J. P. (Eds.). (2013). *Teaching mathematical modelling: connecting to research and practice*. Springer. https://doi.org/10.1007/978-94-007-6540-5.

Strohmaier, A. R., Reinhold, F., Hofer, S., Berkowitz, M., Vogel-Heuser, B., & Reiss, K. (2021). Different complex word problems require different combinations of cognitive skills. *Educational Studies in Mathematics*, *109*(1), 1. https://doi.org/10.1007/s10649-021-10079-4.

Strohmaier, A. R., Ehmke, T., Härtig, H., & Leiss, D. (2023). On the role of linguistic features for comprehension and learning from STEM texts. A meta-analysis. *Educational Research Review*, *39*, 100533. https://doi.org/10.1016/j.edurev.2023.100533.

Sweller, J. (1989). Cognitive technology: some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, *81*(4), 457.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5.
Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, *8*(4), 351–362. https://doi.org/10.1207/s1532690xci0804_5.
Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5.
Taylor, J. A., & McDonald, C. (2007). Writing in groups as a tool for non-routine problem solving in first year university mathematics. *International Journal of Mathematical Education in Science and Technology*, *38*(5), 639–655.
Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 158–179). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.043.
Threlfall, J. (2002). Flexible mental calculation. *Educational Studies in Mathematics*, *50*(1), 29–47. https://doi.org/10.1023/A:1020572803437.
Trabasso, T., & Bouchard, E. (2002). Teaching readers how to comprehend text strategically. In *Comprehension instruction: research-based best practices* (pp. 176–200).
Vajjala, S., & Meurers, D. (2012). *On improving the accuracy of readability classification using insights from second language acquisition*
Van Den Heuvel-Panhuizen, M. (2003). The didactical use of models in realistic mathematics education: an example from a longitudinal trajectory on percentage. *Educational Studies in Mathematics*, *54*(1), 1. https://doi.org/10.1023/B:EDUC.0000005212.03219.dc.
Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of world problems*. Swets & Zeitlinger B.V..
Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: a survey. *ZDM*, *52*(1), 1–16. https://doi.org/10.1007/s11858-020-01130-4.
Vilenius-Tuohimaa, P. M., Aunola, K., & Nurmi, J. E. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology*, *28*(4), 409–426. https://doi.org/10.1080/01443410701708228.
Vos, P. (2020). Task contexts in dutch mathematics education. In M. Van den Heuvel-Panhuizen (Ed.), *National reflections on the netherlands didactics of mathematics: Teaching and Learning in the context of realistic mathematics education* (pp. 31–53). Springer. https://doi.org/10.1007/978-3-030-33824-4_3.
Voyer, D. (2011). Performance in mathematical problem solving as a function of comprehension and arithmetic skills. *International Journal of Science and Mathematics Education*, *9*(5), 5. https://doi.org/10.1007/s10763-010-9239-y.
Wagner, W., Helmke, A., & Rösner, E. (2011). DESI – Eingangsbefragung (Skalenkollektion) (DESI—Initial survey (scale collection)). In *Deutsch Englisch Schülerleistungen International*. https://doi.org/10.7477/6:9:1.
Walkington, C., Clinton, V., & Shivraj, P. (2018). How readability factors are differentially associated with performance for students of different backgrounds when solving mathematics word problems. *American Educational Research Journal*, *55*(2), 362–414. https://doi.org/10.3102/0002831217737028.
Walkington, C., Clinton, V., & Sparks, A. (2019). The effect of language modification of mathematics story problems on problem-solving in online homework. *Instructional Science*, *47*(5), 5. https://doi.org/10.1007/s11251-019-09481-6.
Weinstein, C. E., & Mayer, R. E. (1986). *The teaching cf learning strategies*. Vol. 13.
Werning, R., Löser, J. M., & Urban, M. (2008). Cultural and social diversity: an analysis of minority groups in German schools. *Journal of Special Education*, *42*(1), 1. https://doi.org/10.1177/0022466907313609.
Wienecke, L.-M., Leiss, D., & Ehmke, T. (2023). Taking notes as a strategy for solving reality-based tasks in mathematics. *International Electronic Journal of Mathematics Education*, *18*(3), em744. https://doi.org/10.29333/iejme/13312.
Wijaya, A., van den Heuvel-Panhuizen, M., Doorman, M., & Robitzsch, A. (2014). Difficulties in solving context-based PISA mathematics tasks: an analysis of students' errors. *Mathematics Enthusiast*, *11*(3), 3. https://doi.org/10.54870/1551-3440.1317.
Zaitoun, E., Shana, Z., Shater, A., Naser, K., & Mukattash, Z. (2023). Does flipping the classroom with videos and notetaking promote high school students' performance in mathematics? *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(6), em2274. https://doi.org/10.29333/ejmste/13200.