Learning and Instruction 101 (2025) 102235

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

The relation between text comprehension and computation in mathematical word-problem solving

Anselm R. Strohmaier [a,b,*], Christian Schons [b], Alina Knabbe [c], Markus Vogel [d], Kyra Saado [b], Andreas Obersteiner [b]

[a] University of Education Freiburg, Institute of Mathematics Education, Freiburg, Germany
[b] Technical University of Munich, TUM School of Social Sciences and Technology, Department of Educational Sciences, Munich, Germany
[c] Leuphana University of Lueneburg, Institute of Mathematics and Its Didactics, Lueneburg, Germany
[d] University of Education Heidelberg, Institute of Mathematics, Heidelberg, Germany

ARTICLE INFO

Keywords:
Word-problem solving
Eye tracking
Text comprehension
Mathematics

STRUCTURED ABSTRACT

Background: Mathematical word-problem solving involves both text comprehension and computation, but existing process models differ in their assumptions about whether these processes occur linearly or flexibly, and whether they are situated in integrated or separable mental representations.
Aims: Using eye-tracking, we investigated how the mathematical and linguistic problem difficulty and abilities of the solver and the position of the question affected the solution process, using three categories of effects that can be explained by some models, but not the others.
Sample: Participants were 50 undergraduate students.
Methods: Participants solved 16 three-line traditional mathematical word problems that were independently varied by numerical, lexical, and syntactic problem difficulty and by the position of the question in a within-subject design. Abilities were assessed with standardized instruments. By using eye tracking, visual attention on text and numbers was distinguished in two solution phases and analyzed using linear mixed models.
Results: Linguistic problem difficulty and abilities affected attention on text in the first solution phase, and mathematical problem difficulty and abilities affected attention on numbers and text in the second solution phase.
Putting the question first lead to less attention on numbers and text during the first solution phase, but more attention in the second phase.
Conclusions: Our results concur with models that assume text comprehension precedes computation but occurs in a shared mental representation. However, findings regarding the question position indicate that some mathematical processes occur already in the first solution phase.

1. Introduction

Mathematical word problems are mathematical tasks in which relevant information is presented verbally rather than in mathematical notation (Boonen et al., 2016; Verschaffel et al., 2000). Word problems are embedded in a realistic context to some extent, though the line between word problems and real-world *modelling problems* (e.g., Leiss et al., 2010, see Appendix A) is not clear-cut. Typically, a *traditional word problem* can be solved by the application of one or more mathematical operations to numerical data that are given in the contextualized problem statement (Verschaffel et al., 2000). An example of such a traditional word problem is *"Joe has 3 marbles. Tom has 5 marbles. How many marbles do they have altogether?"* (Riley et al., 1983, p. 160).

Word-problem solving can be considered a bridge between formal mathematics and mathematical problem-solving in the real world and is therefore regarded as an important part of mathematics education (Verschaffel et al., 2000). Because of their verbal and contextualized form, solving word problems requires a combination of linguistic and mathematical thinking (e.g., Daroczy et al., 2015; Daroczy, Meurers, et al., 2020; Jaffe & Bolger, 2023).

To understand and support word-problem solving, decades of research have proposed and investigated different cognitive models of

* Corresponding author. University of Education Freiburg, Institute of Mathematics Education, Kunzenweg 21, 79117, Freiburg, Germany.
E-mail address: anselm.strohmaier@ph-freiburg.de (A.R. Strohmaier).

https://doi.org/10.1016/j.learninstruc.2025.102235
Received 6 September 2024; Received in revised form 20 June 2025; Accepted 24 September 2025
Available online 11 October 2025

word-problem solving (referred to here as *process models*). These process models differ regarding the assumed interactions of mathematical and linguistic processes and the mental representations involved (Daroczy, Meurers, et al., 2020; Roth et al., 2025). However, it remains unclear which existing model best describes the solution process for mathematical word problems.

In this study, we analyzed the solution process of traditional word problems with undergraduate students through the observation of eye movements. We investigated how problem difficulty and learner abilities affect the solution process and how these effects align with the different process models.

1.1. Process models of word-problem solving

Because word problems are in essence mathematical tasks in natural language, their solution requires a process that transforms the written problem into a mental representation of the problem situation described in the text (we refer to this process as *text comprehension*; McNamara & Magliano, 2009), and a computational process to obtain a numerical answer by applying the correct operation to the given numerical values (referred to here as *computation*; Paige & Simon, 1966). All process models of word-problem solving include text comprehension and computation, but they differ in (A) the suggested order of the two processes and (B) whether they assume that the two processes share one mental representation. Regarding A, we distinguish a *linear* and a *flexible* process order; regarding B, we distinguish *separable* and *integrated* mental representations. This distinction is described in detail in the following sections and is used to distinguish four types of process models. A schematic overview of the four types is given in Fig. 1.

1.1.1. Linear process order with separable mental representations

Early process models of word-problem solving assumed that text comprehension usually precedes computation in a distinct, non-mathematical phase called *translation*, *text processing*, or *first phase*. In this study, we refer to this as *first solution phase*. This phase is typically assumed to end after the first complete reading of the whole problem (Hegarty et al., 1992; Mayer et al., 1984), or after reading the problem description, but before reading the question (Verschaffel et al., 1992). Computations are then performed in a later phase called *execution*, *calculation*, or *second phase*, which we refer to as *second solution phase* (Daroczy et al., 2015; De Corte & Verschaffel, 1985; Jaffe & Bolger, 2023; Kintsch & Greeno, 1985; Mayer et al., 1984; Reusser, 1990).

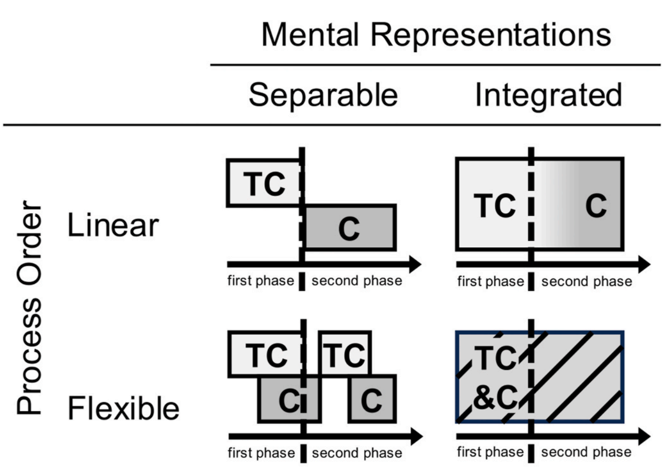

Fig. 1. Schematic representation of four types of process models of mathematical word-problem solving, distinguished by process order and mental representations. The arrow represents time during the solution process. Note that the schema for flexible and separable models reflects one example solution process but the process order can vary between solutions. TC = text comprehension, C = computation.

These early process models also assumed that there are two separable mental representations during word-problem solving: The first one is a mental representation of the text, which is referred to as the *text base*, the *propositional model*, or, in some models, the *situation model*.[1] The second is a mental representation of the schematic mathematical task, which is referred to as *problem model* or *mathematical model* (Jaffe & Bolger, 2023; Kintsch & Greeno, 1985; Nathan et al., 1992). Text comprehension primarily affects the former, while computation is done in the latter representation (Mayer et al., 1984). Because of the sequential order of solution processes and the distinct mental representations, we refer to these process models as *linear* (or *sequential* or *serial*; Roth et al., 2025; Daroczy, Meurers, et al., 2020) regarding the process order, and *separable* regarding the mental representations.

1.1.2. Flexible process order with separable mental representations

In contrast to a strictly linear solution, some process models assume that students alternate between text comprehension and computation or perform both processes simultaneously throughout the solution process. For example, this is the case for circular models in the context of *mathematical modelling* (Borromeo Ferri, 2006; Leiss et al., 2010), where solvers can repeatedly switch back and forth between text comprehension and computation during both the first and second solution phase. This can be generalized to traditional word-problem solving, even though it is particularly plausible for long, real-world modelling problems (Verschaffel et al., 2020).[2] Other models assume that text comprehension and computation occur simultaneously (Roth et al., 2025). In general, when process models assume that there are distinct mental representations but no linear process order, we refer to them as *flexible* (or *parallel* or *simultaneous*; Roth et al., 2025) and *separable*.

1.1.3. Linear process order with integrated mental representations

Some authors challenged the idea of a distinct mathematical mental representation in the form of a rigid solution schema and argued that a single, more flexible mental representation (referred to as *problem model* or *mental model*) is constructed during word-problem solving (Johnson-Laird, 1983; Thevenot, 2010; Thevenot et al., 2007). This problem model is first constructed during text comprehension in the first solution phase and gradually reduced or transformed into a solvable mathematical task during the second solution phase. This unitary mental representation is typically assumed to include mathematical operations as well as aspects of the context and situation at some point during the solution (Daroczy, Meurers, et al., 2020; Gros et al., 2020; Múñez et al., 2013; Staub & Reusser, 1995). We refer to such process models that assume a linear process order but an integrated mental representation as *linear* and *integrated*.

1.1.4. Flexible process order with integrated mental representations

Some models do not strictly separate text comprehension and computation processes (e.g., De Corte et al., 1990; De Corte & Verschaffel, 1986a). For example, Bergqvist and Österholm (2010) suggested an integrated model of word-problem solving based on the assumption that cognitive processes of text comprehension and computation are sometimes indistinguishable. Consequently, there might be only one mental representation, which continuously includes all linguistic and mathematical processes during both solution phases (Bergqvist & Österholm, 2010; Gros et al., 2020). We refer to these

[1] The term *situation model* is used with slightly different purposes and meanings in models of word-problem solving. In general, the situation model refers to any situational representation of the word problem, but Thevenot (2007, 2010) use the term more specifically in contrast to schema-based mental representations.

[2] Note that in Fig. 1, circular models are not depicted as circular but unwrapped.

models as *flexible* and *integrated*.

There is another model that is not included in Fig. 1: Under certain circumstances, students will ignore the context of a word problem and apply a *direct translation strategy* (Verschaffel et al., 2000). This refers to directly inferring the mathematical solution strategy without considering situational factors (Hegarty et al., 1992; Paige & Simon, 1966), for example, by choosing the operation based on a keyword (e.g., to always use an addition when the word "more" is present). The direct translation strategy can sometimes lead to a correct solution, but it is generally unreliable. Thus, it is considered an undesirable strategy and is expected to occur mostly in low-achieving solvers (Jaffe & Bolger, 2023; Paige & Simon, 1966; Verschaffel et al., 2000). Therefore, we do not include it as a process model here.

1.2. Attentional effects on the process of mathematical word-problem solving

The differences between the four types of process models regarding process order and mental representations lead to different assumptions regarding how different parts of the word problem are processed during the first and second solution phase. By systematically and independently varying characteristics of word problems and comparing solvers of differing abilities and by observing how these differences affect the solution process, it is possible to judge the plausibility of different process models of word-problem solving.

For the purpose of this study, we describe three categories of effects on the cognitive processing of different parts of word problems, depending on linguistic and mathematical problem difficulty, the position of the question, and mathematical and linguistic abilities of the solver (see Fig. 2).

For this categorization, we refer to:

- *mathematical problem difficulty* as the difficulty of the underlying symbolic mathematical task of a word problem, for example, regarding the complexity of the operation, the numerical magnitude, or required manipulations to numbers (e.g., carry and borrow; Daroczy et al., 2015; Roth et al., 2025),
- *linguistic problem difficulty* as the difficulty of the language of the word problem, irrespective of the mathematical content. In word problems, this could be lexical complexity, which is assumed to affect local decoding of words and can be influenced by word frequency, word length, compounds, word semantics, or nominalizations (Jucks & Paus, 2012; Strohmaier et al., 2023), and syntactic problem difficulty, which is assumed to influence the construction of a mental representation and is affected by sentence length, the semantic relations, subordinate clauses, or passive voice (Daroczy et al., 2015; Jaffe & Bolger, 2023; Owens, 2016; Strohmaier et al., 2023),
- *mathematical abilities* as any abilities affecting the solution of the underlying symbolic mathematical task of a word problem, for example, arithmetic abilities, and
- *linguistic abilities* as any abilities affecting the comprehension of the language of the word problem, irrespective of the mathematical content, for example, word recognition, vocabulary knowledge or reading fluency.

Furthermore, we distinguish between the cognitive processing of two specific parts of the word problem during these two phases: We refer to:

- *situational text* as all parts of the word problem that are not required for inferring and solving the correct computation[3] and to
- *numbers* as all numerical information in the word problem.

We operationalize cognitive processing as visual attention, where more visual attention indicates higher cognitive processing (Carrasco, 2011). Based on this, we derive the following three categories of effects:

- **Canonical effects:** Some effects are predicted by all types of process models. Thus, we refer to these effects as *canonical effects*: All models have in common that the first phase includes (but is not necessarily limited to) text comprehension, and the second phase includes (but is not necessarily limited to) computation. Accordingly, all models predict that higher linguistic problem difficulty and lower linguistic abilities of the solver should lead to more attention on situational text during the first phase, as its comprehension requires more processing effort (e.g., Böswald, 2024; Daroczy et al., 2025; Dröse et al., 2021; Leiss et al., 2019). Vice versa, higher mathematical problem difficulty and lower mathematical abilities should lead to more attention on numbers during the second solution phase (e.g., Hegarty et al., 1992; Hegarty et al., 1995), since computation requires more effort. Moreover, all models predict that placing the question at the beginning of the problem instead of the end will support text comprehension during the first phase by specifying the reading goal and helping solvers to identify relevant information (Böswald & Schukajlow, 2022). This should lead to less attention on situational text during the first phase when the question is placed first.
- **Phase-crossover effects:** We refer to *phase-crossover effects* when individual abilities and problem difficulty have effects in the phase opposite to the one in which their canonical effects are expected. Flexible models predict that text comprehension also occurs during the second phase, and computation also occurs during the first phase. Therefore, higher linguistic problem difficulty and lower linguistic abilities should also increase attention on situational text during the second solution phase, and higher mathematical problem difficulty and lower mathematical abilities should also lead to more attention on numbers during the first solution phase. No process model predicts that the question position should affect the second solution phase.
- **Attention-crossover effects:** We refer to *attention-crossover* effects when individual abilities and problem difficulty affect attention on the word problem part opposite to the one in which canonical effects are expected. Integrated models predict attention-crossover effects: Because situational and numerical information is stored in the same mental representation, cognitive resources are shared between processing numbers and situational text (Daroczy et al., 2015; Ott et al., 2018). Thus, if either text comprehension or computation requires more resources due to higher problem difficulty or lower abilities, this should increase attention on numbers and situational text alike. Regarding the position of the question, both flexible and integrated models predict attention-crossover effects: In integrated models, identifying relevant information by reading the question first should free up resources in the joint mental representation and allow for quicker and more goal-directed processing of numbers, leading to less attention on numbers (Bergqvist & Österholm, 2010; Böswald & Schukajlow, 2022; Thevenot et al., 2004). In contrast, according to flexible models, placing the question first should initiate computation earlier during the first solution phase, leading to more attention on numbers during the first solution phase and less attention during the second solution phase (Devidal et al., 1997; Thevenot et al.,

[3] In this study, we defined situational text as all parts of the problem that would not make the problem unsolvable when deleted. Note that situational text therefore does not include all non-numerical parts of the word problem; some relevant areas are neither numbers nor situational text.

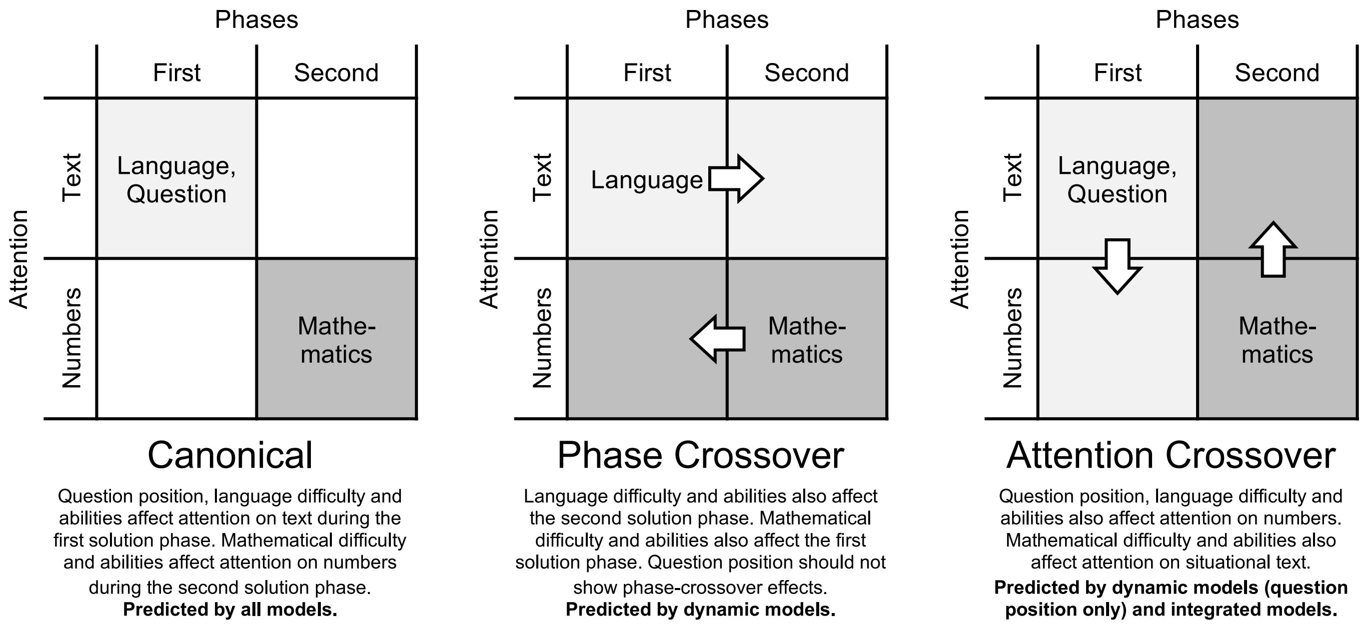

Fig. 2. Three categories of effects that describe how question position, problem difficulty and solver characteristics affect solution phases and attention.

2007). Notably, flexible and integrated models therefore predict opposite attention-crossover-effects, which might also cancel each other out.

In the following, we review the existing literature based on these categories, particularly focusing on evidence for any of the two crossover effects.

1.2.1. Effects of mathematical problem difficulty and abilities

Existing studies on the mathematical problem difficulty of word problems and mathematical abilities can be interpreted with regard to our categorization. Findings regarding phase crossover effects are mixed: Böswald (2024) focused on the comprehension of modelling problems by secondary school students and reported that reading abilities, but not mathematical abilities affected the duration of first reading, finding no crossover effect. In favor of crossover effects, Thevenot et al. (2007) found that fourth graders' reading times in the first solution phase where longer for Compare 1 problems than for Combine 1 and Compare 2 problems (Riley et al., 1983). De Corte and Verschaffel (1986a, 1986b) reported that word-problem difficulty and students' combined verbal and numerical abilities also affected re-reading in the first solution phase for traditional, Riley-type word problems (Riley et al., 1983). However, their descriptive findings stemmed from only six first graders, and word-problem difficulty depended not only on mathematical, but also linguistic problem features. Recently, Roth et al. (2025) showed that number processing by adults was affected by mathematical problem difficulty in traditional arithmetic word problems even if they were not solvable due to missing information. The authors interpret this as evidence that some processing of numbers does occur simultaneously with text comprehension during the first phase, since no computation is required once students have finished comprehending the text. Cook and Rieser (2005) showed that some fifth graders discriminated more flexibly between relevant and added irrelevant numerical information in traditional word problems during the first solution phase. They hypothesize that this might be due to learning effects after several problems of the same structure. Leiss et al. (2019) found that mathematical abilities influenced the comprehension phase and situation model quality in modelling problems for secondary school students, but this might not apply to shorter word problems. Regarding attention crossover effects, (Daroczy et al., 2025) reported for secondary school students that mathematical operations influenced attention on non-numerical areas of traditional Change problems, but these areas were not limited to situational text. In sum, there are findings that point towards crossover effects, but the evidence is not fully clear.

1.2.2. Effects of linguistic problem difficulty and abilities

Compared to mathematical problem difficulty and abilities, fewer studies addressed crossover effects of linguistic problem difficulty and abilities. Leiss et al. (2019) reported a phase crossover effect for modelling problems, such that text comprehension extended to late modelling stages. Attention crossover effects were reported for second graders by De Corte et al. (1990) for Riley-type word problems (Riley et al., 1983), such that semantic complexity led to longer fixations on numbers and words in the second solution phase, but a limitation in this study was that semantic complexity also affected the mathematical problem difficulty. Verschaffel et al. (1992) found that the consistency between the relational term and the mathematical operation in one-step addition and subtraction problems affected reading in the first phase for third graders and university students, indicating a phase crossover effect from the mathematical structure in the text comprehension phase.

1.2.3. Question position

Studies on the question position mostly reported canonical effects, such that placing the question first led to quicker text comprehension (e.g., Böswald, 2024; Thevenot et al., 2007). Thevenot et al. (2004) indirectly investigated a phase-crossover effect by examining whether numbers are maintained in undergraduate students' working memory during a line-by-line presentation when the question is placed first, or if they are already transformed in a computational process. Their results suggested a phase-crossover effect for one type of combine-and-compare problem in which intermediate results could be obtained, but not for a pure combine problem and not for problems in which the last presented quantity was required for the first calculation step. This would indicate that calculations can sometimes occur in the first phase when the question is placed first. To our knowledge crossover effects of the question position have not been explicitly investigated.

2. The present study

It is still discussed among researchers how language and mathematics interact in the solution process of mathematical word problems (Daroczy, Meurers, et al., 2020). Yet the studies that made use of process data to investigate how the solution process is affected by the linguistic and mathematical complexity of the problem, the abilities of the solver, and the position of the question share several limitations. These include possible contamination between linguistic and mathematical problem variations, no distinction between solution phases, assessment of mathematical and linguistic abilities with non-standardized instruments, the use of overly simple prototype word problems, or the use of a data collection methodology that might interfere with the solution process.

In the present study, we addressed these research gaps, combining approaches from previous research to investigate which type of process model fits the solution process of challenging one-step word problems in undergraduate students best. To that end, we systematically and independently varied mathematical and linguistic problem difficulty and the question position and used standardized measures of solvers' linguistic and mathematical abilities. To distinguish between reading phases and to assess visual attention as an indicator of cognitive processing on numbers and situational text, we recorded eye movements during the solution process, which are assumed to be closely related to cognitive processing during text comprehension (Just & Carpenter, 1980; Rayner et al., 2012) and during word-problem solving (Strohmaier, 2020). We posed the following research question:

How do linguistic and mathematical problem difficulty, language and mathematical abilities, and the position of the question affect fixation durations on numbers and situational text in the first and second solution phase of mathematical word-problem solving?

2.1. *Hypothesis 1: canonical effects*

All existing models on the process of word problem predict that higher linguistic problem difficulty and lower linguistic abilities lead to longer fixation durations on situational text during the first solution phase, and that higher numerical problem difficulty and lower mathematical abilities lead to longer fixation durations on numbers during the second solution phase. Moreover, all models predict that placing the question at the beginning should lead to shorter fixation durations on situational text during the first solution phase.

2.2. *Hypothesis 2: crossover effect predicted by flexible models*

According to flexible models, phase crossover-effects should occur: Higher numerical problem difficulty and lower mathematical abilities should also lead to longer fixation durations on numbers during the first solution phase; higher linguistic problem difficulty and lower linguistic abilities should also lead to longer fixation durations on situational text during the second solution phase. Flexible models also predict an attention-crossover effect for question position, such that placing the question at the beginning of the word problem should lead to longer fixation durations on numbers during the first solution phase.

2.3. *Hypothesis 3: crossover effect predicted by integrated models*

Integrated models predict attention-crossover effects, such that higher numerical problem difficulty and lower mathematical abilities should also lead to longer fixation durations on situational text, and higher linguistic problem difficulty and lower linguistic abilities should also lead to longer fixation durations on numbers. Moreover, placing the question first should lead to shorter fixation durations on numbers during first reading.

3. Methods

3.1. *Participants*

We applied a simulation-based a-priori sample size estimation for linear mixed models as described by Kumle et al. (2021) for a power of β*-1* = 0.8 and expected effect sizes of $d > 0.20$ for word problem manipulations and $r > 0.40$ for abilities, plus an expected data loss of 10 %. This resulted in a minimum sample size of $n = 48$ (see Appendix D).

We recruited 50 undergraduate and master's students through advertisement in lectures, the majority (88 %) from mathematics teacher programs (age: $M = 24.8$, $SD = 4.6$, 22 female, 28 male). All participants were fluent in the language of the experiment (German), 7 reported that they had been diagnosed with language difficulties or dyslexia in the past. An ethics approval was not required by institutional guidelines or national regulations.

After data collection, two participants were excluded due to data quality (accuracy $>1°$). One participant was excluded due to mean reaction time of more than 3 *SD* above the average. Two solutions were excluded because of an error in the rotation design, and 3 solutions were excluded due to a numerical error in the problem. Remaining data were 746 word-problem solutions from 47 participants, which amounts to a total data loss of 6.75 %.

3.2. *Design*

3.2.1. *Word problems*

Sixteen original mathematical word problems were created by the authors consisting of two lines of problem description and one line with a question (see Fig. 3 for an example). To cover a wider range of mathematical content, we constructed four problems each from the content areas of percentages, exponential growth, combinatorics, and probability. All 16 problems were framed within different contexts, such as the number of available seats in a football stadium, interest rates, or tiling a floor. The average problem length was $M = 25.1$ words ($SD = 4.5$ words, minimum = 18 words, maximum = 33 words).

Each word problem was systematically and independently varied along four dimensions. Each dimension had two versions, resulting in a total of 2x2x2x2 = 16 different versions for each word problem context.

Mathematical problem difficulty was varied by changing the numerical problem difficulty. This was done by changing exactly two numerical values to make the computation more challenging, for example, 15 % of 400€ instead of 20 % of 500€. Solution rates per word problem ranged from 39 % to 100 % in the numerically easy versions ($M = 77$ %, $SD = 20$ %) and from 13 % to 91 % in the numerically difficult versions ($M = 57$ %, $SD = 21$ %), $t(15) = 3.63$, $p = .002$, $d = 0.90$.

Linguistic problem difficulty was varied across the two independent dimensions: For lexical problem difficulty, two words were replaced with less frequent synonyms, which increases linguistic difficulty (Jucks & Paus, 2012). Word frequency was determined using the DECOW14ax corpus (Schäfer & Bildhauer, 2012). On average, the more frequent words were 37 times more frequent than their less frequent counterparts. Regarding syntactic problem difficulty, the problem description of the easier version consisted of two main clauses in the active voice. In the difficult version, one sentence was changed to a dependent clause, which in German also involves an inversion in word order. Additionally, passive voice was used in the main clause. These changes increase linguistic difficulty (Owens, 2016). On average, problem length increased by $M = 0.4$ words (minimum = −2 words, maximum = +2 words) for higher syntactic problem difficulty.

Finally, we varied the position of the question: For each word problem, a version was created where the question was moved from the third line to the first line of the word problem. The question was the same in both versions.

From these total 256 word problems, 16 distinct sets of problems were constructed. Each set contained one problem of each context, and

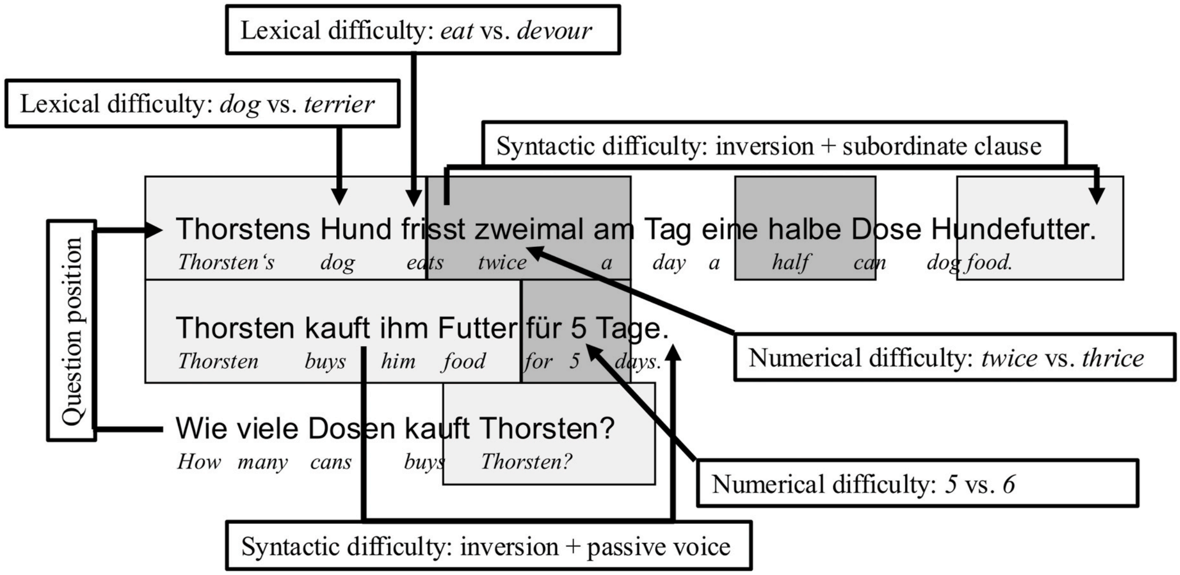

Fig. 3. Example word problem in the easiest version with an English verbatim translation in italics. Marked with arrows are the changes in numerical, lexical and syntactic problem difficulty and the question position. Light grey boxes are the AOI *situational text*, dark grey boxes are the AOI *numbers* (see Section 3.5.2 for AOI definitions).

one problem of each variation. Within each set, the order of the problems was randomized. An example problem and a translation in all 16 versions is given in Appendix B.

3.2.2. Abilities

Mathematical abilities were assessed with 20 self-developed, non-verbal arithmetic tasks that had to be solved on paper without a calculator with a time limit of 7 min. The tasks ranged from arithmetic (e.g., $144 : 4 = x$) to equations (e.g., $4^{(z-3)} = 64$), with natural numbers as solutions, and the test had a maximum score of 20. The test was structurally similar to the module *RE* from the *I-S-T 2000R* (Liepmann et al., 2007). Linguistic abilities were assessed with a timed c-test, consisting of four cloze-tests with 25 blanks each, with a total maximum score of 100 (Grotjahn et al., 2010; Heine, 2017). Each text had a time limit of between 60 and 120 s. Reliability was good in both tests, $\alpha = .80$ (mathematics) and $\alpha = .95$ (language).

3.3. Apparatus

Eye-Tracking data was collected using a Tobii Spectrum with a sampling rate of 600Hz. Word problems were presented on an LCD monitor with a resolution of 1920x1080px and a display area of 529 mm × 298 mm at a distance of 550 mm–650 mm. Word problems were presented in Arial with 24 pt font size, corresponding to 32px line height. The distance between base lines was 108px. In this configuration, one degree of visual angle corresponds to around 40px on the screen.

3.4. Procedure

Participants took part in individual sessions in a quiet laboratory with artificial lighting. Experiments were conducted by one of three trained research assistants. Prior to the experiment, participants were informed about the procedure and gave written informed consent. Abilities and demographic data were assessed on paper. After a short break, participants were made familiar with the eye tracker, which was then calibrated. Each participant solved 16 word problems (one of each context and one of each version) without time limit. Answers were given orally and recorded. Participants were not allowed to take notes during this process. Participants then solved a second set of tasks, which was not part of the study reported here. After the experiment, students received 15€ in cash, where thanked and debriefed. Overall, testing sessions lasted about 45–60 min.

Data was recorded using *Tobii Pro Lab*. A 9-point calibration and 4-point validation was repeated until an accuracy of less than 0.6° and a precision (*SD*) of less than 0.2° was reached. If this was not achieved in the third try, values of less than 1.0° (accuracy) and 0.8° (precision) were accepted. Average achieved accuracy was $M = 0.47°$, average precision was $M = 0.16°$, the average number of calibration attempts was $M = 2.09$.

3.5. Analyses

3.5.1. Timing

The full solution process was defined as the time between the start of the stimulus presentation and the audible onset of participants' final answer, irrespective of whether they gave their answer a whole sentence or a number. Solution phases were distinguished based on eye movement data (see Appendix E). In short, the first solution phase included the process until the word problem was fully read for the first time, i.e., all information was perceived visually. The second solution phase was the remaining time of the full solution process.

3.5.2. Data and AOIs

The Tobii built-in I-VT event detection algorithm was used for fixation detection (threshold 30°/s, minimum fixation duration 80ms). Several build-in filters from *Tobii Pro Lab* were applied (moving median noise reduction, merging of adjacent fixations, interval velocity calculation). AOIs were defined surrounding numbers (or number words), and situational text, with 42px margins to the top and bottom (half of the leading, see Fig. 3 for an illustration). For numbers, 50px margins to the left and right were used. Situational text was defined as 50px away from the closest relevant information. None of the AOIs overlapped. Sensitivity analyses for horizontal margins down to 0px were conducted to ensure all results were robust against smaller margins.

Fixation duration was used as an outcome measure, representing visual attention, which in turn was interpreted as processing effort (Rayner et al., 2012; Strohmaier et al., 2020). Fixation duration was the sum of the duration of all fixations within the respective phase and AOI.

3.5.3. Statistical analyses

Linear Mixed Models were used for analyses, with random effects for participant and problem context. Word problem manipulations were included as dummy variables, abilities were z-standardized and included as continuous predictors. The interaction effect between language and mathematics abilities was included to account for a possible correlation. To reduce model complexity, no further interaction effects were included.

All analyses were conducted using R packages *lme4* (Bates et al., 2015, ver. 1.1–35.3) and *lmerTest* (Kuznetsova et al., 2017, ver. 3.1–3) power analyses were done using *mixedpower* (Kumle et al., 2021, ver. 0.1.0) and *simr* (Green & MacLeod, 2016, ver. 1.0.7)

4. Results

4.1. Descriptive results

Descriptive results for all outcome measures and solution rates are given in Table 1. The appendix includes further analyses on the correlation between the person-level parameters (Table C.1) and effects on problem solution (Table C.2). Of the word problem variations, only numerical problem difficulty significantly affected solution probability, OR = 0.30, $p < .001$. Question position showed a trend such that word problems with the question at the beginning were solved correctly less often, OR = 0.71, $p = .056$. Both mathematical and linguistic abilities were significantly associated with higher solution probabilities, with no significant interaction, OR = 1.67, $p < .001$; OR = 1.35, $p = .008$, OR = 0.90, $p = .379$.

4.2. Effects on fixation durations by phase and area of interest

Table 2 contains the results of four separate linear mixed models, showing effects of all predictors at the word-problem level and person level on fixation durations. Separate analyses were conducted for the first and second solution phase, and for the AOIs *numbers* and *situational text*.

Results show that numerical problem difficulty significantly increased fixation durations both on numbers and situational text in the second solution phase. The effects on fixation durations on numbers were substantially larger than on situational text. Lexical and syntactic problem difficulty led to significantly longer fixation durations on situational text during the first solution phase.

Higher mathematical abilities led to significantly shorter fixation durations both on numbers and situational text in the second phase. Higher language abilities led to a significantly shorter fixation durations on situational text in the first phase.

Placing the question at the beginning of the problem led to a significantly shorter fixation duration on situational text during the first phase. In addition, fixation durations on numbers were significantly shorter in the first phase and significantly longer in the second solution phase when the question was placed at the beginning of the problem. However, these two effects were not consistent for the sensitivity analyses for AOI margins, such that they were not significant when smaller margins were used.

Table 1

Descriptives of outcome variables and fixation durations.

Variable	Mean	*SD*	Median	Min.	Max.
Outcomes					
Solution rate	0.67	0.15	0.69	0.27	0.94
Reaction time	22.69	15.16	17.89	4.54	125.97
Linguistic abilities[1]	68.94	12.79	71	32	89
Mathematical abilities[2]	15.06	3.40	15	6	20
Solution process: Fixation duration [s]					
Full solution process					
Total	18.49	12.28	14.61	3.57	96.71
Numbers	9.12	8.90	6.26	0.14	62.65
Situational text	6.08	4.19	4.99	0.72	33.72
First phase					
Total	6.21	2.63	5.79	1.94	29.04
Numbers	1.35	1.43	1.08	0.00	22.33
Situational text	3.29	1.48	2.97	0.72	9.40
Second phase					
Total	11.73	11.37	8.23	0.00	87.10
Numbers	7.76	8.81	4.78	0.00	61.84
Situational text	2.25	3.13	1.24	0.00	30.93

Note. [1]Range = 0-100. [2]Range = 0-20.

5. Discussion

This study aimed to investigate how linguistic and mathematical problem difficulty, language and mathematical abilities, and the question position affect the solution process of mathematical word-problem solving (see Fig. 4). We used three categories of effects which were predicted differently by different models of word-problem solving (see Section 1.2).

5.1. Canonical effects (Hypothesis 1)

Hypothesis 1 was predicted by all models and fully aligned with our data. When word problems were linguistically more difficult or solvers had lower linguistic abilities, more attention was directed towards situational text during the first solution phase. Empirical research on eye movements during reading of texts shows that that less frequent words and more complex syntactic structures, as well as lower linguistic abilities, increase fixation durations during text comprehension (Rayner et al., 2012). The fact that this was observed during the first solution phase indicates that solvers were indeed comprehending the text at this point. Vice versa, numerical problem difficulty and lower mathematical abilities led to longer fixation durations on numbers during the second solution phase. This indicates that computation happened during this phase. Placing the question at the beginning of the word problem led to shorter fixation durations on situational text during first reading. In line with theory and prior research, this indicates that reading the question at the beginning helped solvers to specify the reading goal and identify relevant information (Böswald, 2024; Böswald & Schukajlow, 2022).

These canonical effects are not a novelty and do not contribute to distinguishing underlying models, but the fact that all canonical effects robustly occurred in our study indicates that our design and methodology were likely suitable to also detect crossover effects. Moreover, they support our premise that under the given circumstances our participants did not or hardly noticeably apply a direct translation strategy, which would not lead to canonical effects.

5.2. Crossover effect predicted by flexible models (Hypothesis 2)

Contrary to the prediction of flexible models, we did not find evidence for phase crossover effects, such that linguistic problem difficulty or abilities would affect the second solution phase, or that mathematical problem difficulty or abilities would affect the first solution phase. The fact that linguistic problem difficulty or abilities had no effect in the second phase indicates that at this point, the linguistic information is integrated into a mental representation in a format that is no longer affected by lexical or syntactic problem difficulty, for example, in a flexible mental model as suggested by Johnson-Laird (1983) and Thevenot (2010). Similarly, the fact that mathematical problem difficulty and abilities did not affect the first solution phase suggests that solvers were not yet computing at this point.

These findings seem to be in contrast with Roth et al. (2025) who reported evidence that mathematical difficulty did affect fixation durations on numbers in non-solvable word problems. While our findings do not support flexible models with computations occurring during the first solution phase, Roth et al. (2025) clearly point towards this direction for their set of tasks and participants. In contrast to our study, Roth

Table 2
Effects of word problem characteristics and person characteristics on fixation durations on situational text and numbers during the first and second solution phase.

Fixed Effects	Situational Text									
	First Phase					Second Phase				
	Estimate	*SE*	95 % CI	*p*	$\Delta R^2_m/\Delta R^2_c$	Estimate	*SE*	95 % CI	*p*	$\Delta R^2_m/\Delta R^2_c$
Intercept	3.01	0.21	[2.59, 3.42]			1.44	0.37	[0.71, 2.17]		
Word Problem Level										
Numerical Difficulty	0.07	0.08	[-0.08, 0.22]	0.389	0.00/0.00	**1.02**	**.20**	**[.62, 1.41]**	**<.001**	**.03/.02**
Lexical Difficulty	**.23**	**.08**	**[.08, .39]**	**.002**	**.01/.01**	−0.27	0.20	[-0.67, 0.12]	0.171	0.00/0.00
Syntactic Difficulty	**.41**	**.08**	**[.26, .56]**	**<.001**	**.02/.02**	0.06	0.20	[-0.33, 0.45]	0.765	0.00/0.00
Question Position[1]	**−.16**	**.08**	**[-.31, -.01]**	**.037**[†]	**.00/.00**	**.78**	**.20**	**[.39, 1.18]**	**<.001**	**.02/.02**
Person Level										
Mathematical Abilities	−0.13	0.09	[-0.32, 0.05]	0.143	0.01/0.00	**−.49**	**.16**	**[-.80, -.18]**	**.002**	**.02/.00**
Linguistic Abilities	**−.51**	**.10**	**[-.70, -.31]**	**<.001**	**.10/.00**	0.15	0.16	[-0.18, 0.47]	0.359	0.00/0.00
Interaction	−0.05	0.10	[-0.25, 0.15]	0.615	0.00/0.00	−0.22	0.17	[-0.55, 0.11]	0.182	0.00/0.00
Full Model										
Omnibus Test				**<.001**	**.15/.51**				**<.001**	**.07/.25**
Random Effects Variance	Estimate	SD				Estimate	SD			
Between Persons	0.33	0.58				0.64	0.80			
Between Content	0.48	0.70				1.19	1.09			
Residual	1.10	1.05				7.49	2.74			
Fixed Effects	Numbers									
	First Phase					Second Phase				
	Estimate	*SE*	95 % CI	*p*	$\Delta R^2_m/\Delta R^2_c$	Estimate	*SE*	95 % CI	*p*	$\Delta R^2_m/\Delta R^2_c$
Intercept	1.52	0.16	[1.22, 1.83]			3.64	1.07	[1.55, 5.74]		
Word Problem Level										
Numerical Difficulty	0.10	0.10	[-0.10, 0.29]	0.331	0.00/0.00	**6.68**	**.49**	**[5.72, 7.64]**	**<.001**	**.14/.15**
Lexical Difficulty	−0.07	0.10	[-0.27, 0.12]	0.455	0.00/0.00	0.09	0.49	[-0.87, 1.05]	0.849	0.00/0.00
Syntactic Difficulty	0.02	0.10	[-0.17, 0.22]	0.810	0.00/0.00	0.39	0.49	[-0.57, 1.35]	0.422	0.00/0.00
Question Position[1]	**−.39**	**.10**	**[-.59, -.20]**	**<.001**	**.02/.02**	**.99**	**.49**	**[.02, 1.95]**	**.044**[†]	**.00/.00**
Person Level										
Mathematical Abilities	−0.08	0.06	[-0.20, 0.04]	0.159	0.00/0.00	**−2.09**	**.51**	**[-3.09, -1.09]**	**<.001**	**.05/.00**
Linguistic Abilities	−0.10	0.06	[-0.22, 0.03]	0.114	0.00/0.00	0.36	0.53	[-0.68, 1.40]	0.480	0.00/0.00
Interaction	−0.03	0.06	[-0.16, 0.09]	0.586	0.00/0.00	−0.75	0.54	[-1.82, 0.31]	0.156	0.01/0.00
Full Model										
Omnibus Test				**.002**	**.03/.14**				**<.001**	**.20/.44**
Random Effects Variance	Estimate	*SD*				Estimate	*SD*			
Between Persons	0.05	0.23				8.94	2.99			
Between Content	0.18	0.42				10.34	3.22			
Residual	1.79	1.34				44.38	6.66			

Note. [1]Question position was coded 0 = question at the end (third line) 1 = question at the beginning (first line). [†]These effects were not robust in sensitivity analyses for AOI margins.

et al. (2025) varied mathematical difficulty by altering required carry and borrow operations in addition and subtraction problems, which might be a more salient problem feature than numerical difficulty. Further, our distinction between the first and second phase was based on the moment when the word problem was fully read for the first time, whereas Roth et al. (2025) assume that students might re-read the problem before deciding that it is non-solvable and give an answer. This re-reading would, in our paradigm, already fall within the second solution phase and might lead to different results. These contrasting findings underscore that the solution process of word-problem solving likely differs depending on the problem and the solver.

5.3. *Crossover effect predicted by integrated models (Hypothesis 3)*

In addition to canonical effects, we found that mathematical problem difficulty and abilities also affected the processing of situational text. This attention crossover effect was predicted by integrated models. If solvers transformed the problem into a purely mathematical model before computation, attention on situational text should not be affected by the mathematical problem difficulty and abilities. The results suggests that the text remains part of the mental representation and is processed simultaneously and with shared cognitive resources. However, it is also possible that students looked at the problem randomly while mentally computing, making the eye-mind assumption less reliable for determining the exact focus of attention during the second phase. No attention-crossover effects were observed for linguistic problem difficulty and abilities.

We found shorter fixation durations on numbers during the first phase when the question was placed first. In line with Thevenot et al. (2004), this would indicate that placing the question at the beginning does not lead to immediate computations as would be predicted by flexible models, since this would likely lead to longer, not shorter, fixation durations on numbers. The findings align better with the assumption of integrated models that numbers are processed together with text in one mental representation, which is constructed quicker and more efficiently when the question is placed first. This means that mathematical processes could play a role during the first solution phase, even if it is not be in the form of early computation (Thevenot et al., 2004, 2007). Such processes could be structuring, associating or retrieval of memorized facts (Bergqvist & Österholm, 2010).

5.4. *Further findings*

Going beyond our hypotheses, we found that solvers directed more attention towards situational text during the second solution phase when the question was placed first, and a slight trend towards more attention on numbers. This might indicate some sort of compensatory effect: While the solution process might initially be accelerated, this

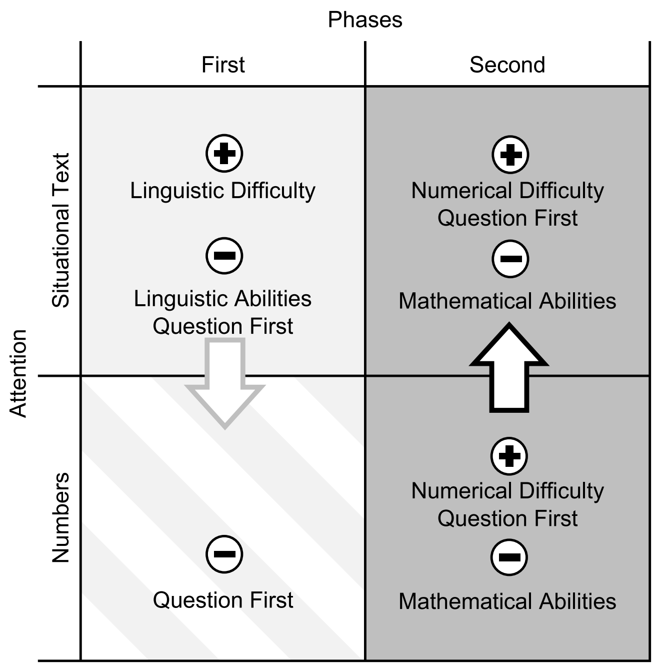

Fig. 4. Observed effects on fixation durations in both phases and on both AOIs. The plus symbol indicates longer fixation durations, the minus symbol indicates shorter fixation durations. Arrows indicate the observed attention crossover effects (see Fig. 1).

seems to lead to detrimental effects in the later stages of the solution process. One explanation could be that reading the question first tempts students to skip supposedly irrelevant information, leading to a reduced situation model which in some cases might later prove to be insufficient (Böswald, 2024). If readers had to re-read aspects of the problem, this might increase fixation durations in the second solution phase.

5.5. Implications

In sum, our findings support linear, integrated models of word-problem solving. These models assume that there is only one flexible mental representation in which text comprehension processes precede computation. We would argue that the *Semantic Congruence Model* (SECO; Gros et al., 2020) reflects an example of such a process model.

This indicates that linguistic and mathematical processes are connected during word-problem solving through a joint mental representation. Consequently, teaching text-comprehension strategies for word-problem solving should be particularly valuable if those strategies are working towards preparing the mathematical model, not merely to "strip away" the story context and identify keywords (Powell et al., 2020). This was also supported by the unexpected finding of a longer second solution phase when the question was placed first: Supporting text comprehension superficially might initially lead to quicker linguistic processing, but might result in an incomplete mental representation that requires more time to transform into an adequate mathematical model.

5.6. Limitations

All models of word-problem solving assume a consistent process across word problem solutions. However, it is also plausible to assume that individual solvers might choose their approach differently from problem to problem, depending on the context, the problem difficulty, or the underlying mathematical structure. We did not take this possibility into account in our analytical design, as we investigated cumulated effects across individuals. Future studies could try to categorize underlying processes on a problem-by-problem level. This would require much more detailed categorizations, for example, by reintroducing human coding of solution processes like De Corte and Verschaffel (1986a, 1986b), or by using machine learning methods to automatically categorize the rich process data from eye tracking or multiple data sources (Strohmaier et al., 2024).

A key limitation is that, more generally, the solution process might differ between the word problem type and between student populations (Bergqvist & Österholm, 2010; Strohmaier et al., 2022). Most of the models of mathematical word-problem solving were developed on the basis of prototype arithmetic word problems (Riley et al., 1983) for which it might make sense to just translate the problem into a mathematical schema, in the way that linear and separable models suggest (Mayer et al., 1984). More complex word problems or modelling problems could require a stronger integration of contextual cues and a more comprehensive situation model (Gros et al., 2020). We used traditional word problems in our study which were based on challenging mathematical structures. Using easier, simple arithmetic problems or more contextualized modelling problems might lead to different results.

Similarly, our participants were undergraduate students, and the solution processes might differ for younger students. It is not unlikely that distinguishable mental representations (in line with separable process models) are more common in younger students, whereas a higher expertise or more training in solving word problems, as well as more straightforward problems, could increase the likelihood that computation is initiated immediately during the first solution phase. Even though the categorization of effects applied in this study can be generalized across word problems, their interpretation requires a careful consideration of these factors.

Finally, some limitations are caused by the eye-tracking technology. The fact that students were not allowed to take notes in order to record the whole solution process meant that the solution might be different from a solution process on paper. Second, while the visual focus typically coincides with the attentional focus during reading, it is possible that participants perceived some parts of the word problem in peripheral vision. This could, in theory, appear as an attention-crossover effect in the data. However, to our knowledge, there is so far no evidence of the use of peripheral vision in mathematics tasks.

5.7. Conclusion

Our findings robustly indicate that text comprehension precedes computation in mathematical word-problem solving. While it seems that these processes do not occur in parallel or quickly alternate, they seem to share a joint mental representation. These insights highlight the intertwined nature of linguistics and mathematics during mathematical word-problem solving but also show that there are clear differences in cognitive processing during the different phases of the solution process.

CRediT authorship contribution statement

Anselm R. Strohmaier: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Christian Schons:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. **Alina Knabbe:** Writing – review & editing, Writing – original draft. **Markus Vogel:** Writing – review & editing, Writing – original draft. **Kyra Saado:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. **Andreas Obersteiner:** Writing – review & editing, Resources.

Declarations of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2025.102235.

Data availability

Data will be made available on request.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergqvist, E., & Österholm, M. (2010). *A theoretical model of the connection between the process of reading and the process of solving mathematical tasks. MADIF 7, the 7th Swedish Mathematics Education Research Seminar, Stockholm, January 26-27, 2010*.

Boonen, A. J. H., de Koning, B. B., Jolles, J., & van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in Psychology, 7*, 191. https://doi.org/10.3389/fpsyg.2016.00191

Borromeo Ferri, R. (2006). Theoretical and empirical differentiations of phases in the modelling process. *ZDM, 38*(2), 86–95. https://doi.org/10.1007/BF02655883

Böswald, V. (2024). *Die Rolle der Position der Fragestellung beim Textverstehen von mathematischen Modellierungsaufgaben*. Springer Spektrum. https://doi.org/10.1007/978-3-658-43675-9

Böswald, V., & Schukajlow, S. (2022). Reading comprehension and modelling problems: Does it matter where the question is placed?. *Twelfth Congress of the European Society for research in mathematics education (CERME12)*. Bozen-Bolzano, Italy https://hal.science/hal-03753483.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Cook, J. L., & Rieser, J. J. (2005). Finding the Critical facts: Children's visual scan Patterns when solving story problems that contain irrelevant information. *Journal of Educational Psychology, 97*(2), 224–234. https://doi.org/10.1037/0022-0663.97.2.224

Daroczy, G., Artemenko, C., Wolska, M., Meurers, D., & Nuerk, H. C. (2025). Are text comprehension and calculation processes in word problem solving sequential or interactive? An eye-tracking study in children. *Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale, 79*(2), 206–211. https://doi.org/10.1037/cep0000366

Daroczy, G., Meurers, D., Heller, J., Wolska, M., & Nürk, H.-C. (2020). The interaction of linguistic and arithmetic factors affects adult performance on arithmetic word problems. *Cognitive Processing, 21*(1), 105–125. https://doi.org/10.1007/s10339-019-00948-5

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 Corte, E., & Verschaffel, L. (1985). Beginning first graders' initial representation of arithmetic word problems. *The Journal of Mathematical Behavior, 4*(1), 3–21.

De Corte, E., & Verschaffel, L. (1986a). *Eye-movement data as access to solution processes of elementary addition and subtraction problems*. San Francisco, USA: Annual Meeting of the American Educational Research Association.

De Corte, E., & Verschaffel, L. (1986b). Eye-movements of first graders during word problem solving. In Univ of London Institue of Education (Ed.), *Proceedings of the 10th Conference of the international Group for the Psychology in mathematics education* (pp. 421–426). PME.

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology, 82*(2), 359–365. https://doi.org/10.1037/0022-0663.82.2.359

Devidal, M., Fayol, M., & Barrouillet, P. (1997). Stratégies de lecture et résolution de problèmes arithmétiques. *L'Année Psychologique, 97*, 9–31.

Dröse, J., Prediger, S., Neugebauer, P., Delucchi Danhier, R., & Mertins, B. (2021). Investigating students' processes of noticing and interpreting syntactic language features in word problem solving through eye-tracking. *International Electronic Journal of Mathematics Education, 16*(1). https://doi.org/10.29333/iejme/9674

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

Gros, H., Thibaut, J.-P., & Sander, E. (2020). Semantic congruence in arithmetic: A new conceptual model for word problem solving. *Educational Psychologist, 55*(2), 69–87. https://doi.org/10.1080/00461520.2019.1691004

Grotjahn, R., Schlak, T., & Aguado, K. (2010). SC-Tests: Messung automatisierter sprachlicher Kompetenzen anhand von C-Tests mit massiver textspezifischer Zeitlimitierung. *The C-Test: Contributions from current research*.

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology, 84*(1), 76–84. https://doi.org/10.1037/0022-0663.87.1.18

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccesful problem solvers. *Journal of Educational Psychology, 87*(1), 18–32. https://doi.org/10.1037/0022-0663.87.1.18

Heine, S. (2017). *Fremd-und Zweitsprachenerfolg und seine Erklärung durch Erwerbsalter, kognitive, affektiv-motivationale und sozio-kulturelle Variablen: Eine empirische Studie*. kassel university press.

Jaffe, J. B., & Bolger, D. J. (2023). Cognitive processes, linguistic factors, and arithmetic word problem success: A review of behavioral studies. *Educational Psychology Review, 35*(4), 105. https://doi.org/10.1007/s10648-023-09821-6

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press. https://hal.science/hal-00702919.

Jucks, R., & Paus, E. (2012). What makes a word difficult? Insights into the mental representation of technical terms. *Metacognition and Learning, 7*(2), 91–111. https://doi.org/10.1007/s11409-011-9084-6

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354. https://doi.org/10.1037/0033-295X.87.4.329

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

Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

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., 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

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-struktur-test 2000 R, 2.* Hogrefe.

Mayer, R. E., Larkin, J. H., & Kadane, J. B. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Ed.), *Advances in the Psychology of human Intelligence, 2* pp. 231–273). Lawrence Erlbaum Associates. https://cir.nii.ac.jp/crid/1571698599027538304.

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

Múñez, D., Orrantia, J., & Rosales, J. (2013). The effect of external representations on compare word problems: Supporting mental model construction. *The Journal of Experimental Education, 81*(3), 337–355. https://doi.org/10.1080/00220973.2012.715095

Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction, 9*(4), 329–389. https://doi.org/10.1207/s1532690xci0904_2

Ott, N., Brunken, R., Vogel, M., & Malone, S. (2018). Multiple symbolic representations: The combination of formula and text supports problem solving in the mathematical field of propositional logic. *Learning and Instruction, 58*, 88–105. https://doi.org/10.1016/j.learninstruc.2018.04.010

Owens, R. E. (2016). *Language development: An introduction* (9th ed.). Pearson.

Paige, J. M., & Simon, H. A. (1966). Cognitive processes in solving algebra word problems. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory* (pp. 51–119). Wiley.

Powell, S. R., Berry, K. A., & Benz, S. A. (2020). Analyzing the word-problem performance and strategies of students experiencing mathematics difficulty. *The Journal of Mathematical Behavior, 58*, Article 100759. https://doi.org/10.1016/j.jmathb.2020.100759

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *Psychology of reading* (2 ed.). Psychology Press.

Reusser, K. (1990). From text to situation to equation: Cognitive simulation of understanding and solving mathematical word problems. In H. Mandl, E. De Corte, N. S. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction in an international context* (pp. 477–498). Pergamon.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. H. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). Academic Press.

Roth, L., Nuerk, H. C., Cramer, F., & Daroczy, G. (2025). Can't help processing numbers with text: Eye-tracking evidence for simultaneous instead of sequential processing of text and numbers in arithmetic word problems. *Psychological Research, 89*, 50. https://doi.org/10.1007/s00426-024-02069-x

Schäfer, R., & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *European language resources association (ELRA)*.

Staub, F. C., & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 285–305). Lawrence Erlbaum Associates, Inc.

Strohmaier, A. R. (2020). *When reading meets mathematics. Using eye movements to analyze complex word problem solving*. Munich: Technical University of. https://doi.org/10.14459/2020md1521471

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*, Article 100533. https://doi.org/10.1016/j.edurev.2023.100533

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye tracking methodology in mathematics education research: A systematic literature review.

Educational Studies in Mathematics, 104, 147–200. https://doi.org/10.1007/s10649-020-09948-1

Strohmaier, A. R., Mora-Ruano, J. G., Schons, C., & Obersteiner, A. (2024). Can a machine learning algorithm Tell right from wrong in eye movements of mathematical word problem solving?. In *Beiträge zum Mathematikunterricht 2024*. WTM.

Strohmaier, A. R., Reinhold, F., Hofer, S., Berkowitz, M., Vogel-Heuser, B., & Reiss, K. (2022). Different complex word problems require different combinations of cognitive skills. *Educational Studies in Mathematics, 109*(1), 89–114. https://doi.org/10.1007/s10649-021-10079-4

Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta Psychologica, 133*(1), 90–95. https://doi.org/10.1016/j.actpsy.2009.10.004

Thevenot, C., Barrouillet, P., & Fayol, M. (2004). Représentation mentale et procédures de résolution de problèmes arithmétiques : l'effet du placement de la question. *L'année psychologique, 104*(4), 683–699. https://doi.org/10.3406/psy.2004.29685

Thevenot, C., Devidal, M., Barrouillet, P., & Fayol, M. (2007). Why does placing the question before an arithmetic word problem improve performance? A situation model account. *Quarterly Journal of Experimental Psychology, 60*(1), 43–56. https://doi.org/10.1080/17470210600587927

Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency Hypothesis. *Journal of Educational Psychology, 84*(1), 85–94. https://doi.org/10.1037/0022-0663.84.1.85

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Swets & Zeitlinger.

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education, 52*(1), 1–16. https://doi.org/10.1007/s11858-020-01130-4