Learning and Instruction 92 (2024) 101902

Contents lists available at ScienceDirect

Learning and Instruction

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

Training flexibility in dealing with additive situations

Laura Gabler [1,*], Stefan Ufer [2]

LMU Munich, Theresienstr. 39, 80333, Munich, Germany

ARTICLE INFO

Keywords:
Additive word problems
Flexibility
Language
Comparison of sets

ABSTRACT

Objective: Students' difficulties with word problems have been the subject of research for decades. Many studies identified students' ability to construct a situation model that reflects the word problem's situation structure correctly as a major factor. To overcome such difficulties, prior works suggested to provide learners with strategies, which comprise to restructure the situation model by integrating different perspectives on the presented situation. Corresponding trainings have not been investigated systematically yet.
Methods: We report on an experimental feasibility study investigating a training targeting the proposed strategies. Students from ten grade 2 classrooms ($N = 115$) in Germany participated in the study. The ten-day training focused on generating and comparing different perspectives on given situations but did not include any word problem solving.
Results: Students participating in the training showed significantly higher progress in their ability to restructure situation models and their word problem solving skills from pre-to follow-up test than students from the control-group. The effect of the training was not influenced by students' language skills.
Conclusion: The results indicate that it is feasible to foster word problem solving skills by solely training how to restructure the initial situation model generated from a word problem.
Practice: Since the experimental group received additional support in contrast to the control group, it is impossible to draw conclusions about the importance of the training for regular mathematics lessons, beyond the fact that the training is effective in principle.
Implications: The approach should be compared to other approaches to foster word problem solving.

1. Introduction

Many studies reported that word problem solving is a particularly challenging task for primary school students (for an overview, see Daroczy et al., 2015). In school, teachers often draw on strategies such as "what I know, what I look for" to support learners in solving difficult word problems (Goulet-Lyle et al., 2020). Students are asked to identify the given sets and derive the solution from this information. However, such strategies do not encourage learners to approach problems flexibly, since they imply a linear, fixed route toward solving word problems and still allow superficial processing (e.g., by keyword strategies). Flexibility is assumed to be essential when dealing with new, unfamiliar situations (Warner et al., 2003). If learners encounter a difficult word problem, they may use flexibility to transfer knowledge and skills from contexts they are already familiar with and use this knowledge or these skills for the solution. Flexible thinking distinguishes good problem solvers from poor problem solvers: According to Schoenfeld (2007, p. 60), good problem solvers are "flexible and resourceful. They have many ways to think about problems – alternative approaches if they get stuck, ways of making progress when they hit roadblocks, of being efficient with (and making use of) what they know." Researchers suggested specific strategies, which encourage such flexibility in the context of word problem solving (Greeno, 1980; Stern, 1993). These strategies build on the idea that learners can generate additional perspectives to a word problem and interpret the described situation as a familiar, more accessible word problem.

We use such strategies as the foundation for a training program that aims at developing flexibility in dealing with the additive situations described in word problems. We investigate, if fostering flexibility in dealing with additive situations by providing the suggested strategies may enhance students' understanding of difficult word problems.

* Corresponding author.
E-mail address: laura.gabler@lmu.de (L. Gabler).
[1] Research interests: word problem solving; role of language in mathematics learning.
[2] Research interests: arithmetic skill development; role of language in mathematics learning; interventions on mathematics conceptual understanding.

https://doi.org/10.1016/j.learninstruc.2024.101902
Received 10 September 2023; Received in revised form 9 January 2024; Accepted 28 February 2024
Available online 20 March 2024

2. Theoretical background

There is a long tradition of national and international research on word problem solving (Daroczy et al., 2015; Kintsch & Greeno, 1985; Stern, 1998). Typical word problems contain verbally described mathematical problems, which can be solved by applying mathematical operations (Verschaffel et al., 2020). Contrary to context-free arithmetic tasks (e.g., “How much is 5 plus 7?”), such word problems describe mathematical operations within real-world phenomena (e.g., “Susi had 5 marbles. Then, she got 7 marbles more. How many marbles does Susi have now?”). From the perspective of mathematics education, traditional word problems need to be distinguished from real-world problems (Verschaffel et al., 2020): Whereas the latter aim at mastering mathematics in authentic, complex everyday situations, word problems in a traditional sense focus on teaching different meanings of mathematical concepts (Stern, 1998). In the classroom, they primarily serve the purpose of addressing various situation types (e.g., increasing a set) that can be described by a mathematical concept (e.g., addition) and their verbal description. As other works in the field, we focus on *additive one-step word problems* (e.g., Huang et al., 2019; Powell et al., 2017; Stern & Lehrndorfer, 1992), which can be solved with a single arithmetic operation (addition or subtraction) and do not contain irrelevant information.

2.1. Word problem solving

A range of process models (e.g., Blum & Leiβ, 2007; Kintsch & Greeno, 1985) describe prototypically how students approach word problems. They assume that learners construct different mental models. This idealized process can be described as “transformational” (Czocher, 2018): Learners transform a real-world problem into a mathematical problem and then transform the solution back again.

Two different types of models play a key role in the solution process (Kintsch & Greeno, 1985; Verschaffel et al., 2020): The *situation model* and the *mathematical model*. Learners encounter a certain *text base*, which is the verbal description of the given situation. Proceeding from this text base, learners construct these two models sequentially. The situation model is the learner’s internal, mental presentation of the given situation (Czocher, 2018) and typically comprises relations among certain quantities from the situation (e.g., part-whole relations). A mathematical model, which is referred to as a “representation that is expressed externally and mathematically” (Czocher, 2018, p. 139), describes these relations with mathematical concepts and symbols. Using mathematical techniques, students obtain a numerical result based on this mathematical model. What constitutes mathematical work in this sense may change with students’ mathematical development. While young (e.g., pre-school) students may model the situation with concrete material, their fingers, or mentally, students from grade 1 and 2 may increasingly activate learned mathematical strategies for addition and subtraction, or recall number facts. Writing down or imagining a symbolic mathematical expression (e.g., $5 + 7 = \ldots$) is not a necessary part of a mathematical model. We speak of a mathematical model, if students establish and use mathematical relations between the numeric quantities entailed in the situation model.

According to the modeling cycle by Blum and Leiβ (2007), interpreting and validating the achieved results are important steps when solving complex, authentic real-world word problems. However, the simplified situations in additive one-step word problems do not contain much potential for interpretation and validation (Kaiser, 2017). Thus, we focus on constructing a situation model and a mathematical model in this work.

Task features influence the difficulty of additive one-step word problems. For example, a *linguistic complexity* of the text base may cause comprehension obstacles and influence the learners’ construction of a situation model (Barbu & Beal, 2010; Plath & Leiss, 2018). A word problem’s difficulty may also be influenced by the *arithmetic complexity* of the numbers used in the presented situation (Daroczy et al., 2020).

Even when reducing the linguistic and arithmetic complexity of a word problem as far as possible, there are still differences in difficulty caused by the *features of the situation structure*: semantic structure, additive or subtractive wording, and unknown set (e.g., Stern, 1998).

2.1.1. Semantic structure

The same mathematical structure (e.g., an additive operation such as $5 + 4 = 9$) can describe different real-world phenomena (Fig. 1). Commonly, these phenomena have been classified into three or four types of additive one-step word problems (so-called “semantic structures”, e.g., Riley et al., 1983).

Additive word problems can describe situations referring to the increase or decrease of a quantity (*change*), the combination of two quantities (*combine*), or the comparison of two quantities (*compare*). *Equalize* problems, a less common type, combine features of change and compare problems. Here, one set is initially compared with a second set (e.g., Susi’s marbles and Max’s marbles). One set is then changed (e.g., adding four marbles to Susi’s set), so that its cardinality is equivalent to the second set. This means that, instead of equalizing both sets (e.g., Max gives marbles to Susi), only one set is changed. While *dynamic* word problems (change, equalize) describe actions, combine and compare problems refer to *static* situations (Riley et al., 1983).

Past research reports relatively consistent findings on the difficulty of the four semantic structures. While change and combine problems are considered rather easy, numerous studies highlighted compare problems as especially difficult semantic structures (e.g., Riley & Greeno, 1988). There are several factors that are discussed to explain the difficulty of compare problems. In compare problems, numbers do not only describe concrete sets, but also the difference between the two concrete sets (Stern, 1993). This difference does not exist as a concrete set, and thus may be harder to represent mentally. Learners can identify a difference set through one-to-one correspondence and counting the excess objects, or through modeling the situation with mathematical operations (Stern, 1998). For the latter, it is crucial to understand addition and subtraction not only as an operation to determine the extent of quantitative change, but also as a way to model additive relations between quantities. Current models on number concept acquisition allocate such semantic structures in later phases of development (under the term "relationality"; for an overview see Hartmann & Fritz, 2021). There is little evidence on the difficulty of equalize problems, since they were not included in the majority of studies or distinguished as a separate semantic structure. Stern (1994) reported solution rates of 96% for first graders. She explained this high solution rate by pointing out that equalize problems only involve concrete sets and no difference sets.

2.1.2. Additive or subtractive wording

Variations of a word problem’s wording can also describe the same mathematical structure. Fuson et al. (1996) distinguish between *additive* and *subtractive wording (a/s wording)*. Linguistically, the relations in compare problems can be expressed by relational terms, such as “more”, “bigger” (additive wording) or “less”, “smaller” (subtractive wording). By varying the a/s wording, different perspectives on the same situation can be emphasized. For instance, “Max has 4 marbles *more* than Susi” can also be expressed with subtractive wording: “Susi has 4 marbles *less* than Max”. Similarly, dynamic word problems can be expressed with action verbs referring to adding (additive wording, e.g., “to get”, “to buy”) or removing a quantity (subtractive wording, e.g., “to give away”, “to sell”). Combine problems do not allow a similar distinction of a/s wording.

2.1.3. Unknown set

One-step word problems involve three sets, of which one is unknown. For compare problems, these sets are called *reference set*, *difference set*, and *compare set* (e.g., Stern, 1993). Their equivalents in dynamic situations are *start set*, *change set*, and *result set*. Combine problems involve

	Static	**Dynamic**
Part-whole	*Combine* Susi has 5 marbles, Max has 4 marbles. How many marbles do they have altogether?	*Change* Susi had 5 marbles. Then, she got 4 marbles more. How many marbles does Susi have now?
Disjoint sets	*Compare* Susi has 5 marbles. Max has 4 marbles more than she has. How many marbles does Max have?	*Equalize* Susi has 5 marbles. If she gets 4 marbles more, she has as many marbles as Max. How many marbles does Max have?

5 + 4 = 9

Fig. 1. Semantic structures describing the same mathematical structure.

one *whole set* with two (disjoint, but exhaustive) *subsets*. Studies have shown that word problems with an unknown reference/start set or unknown subset are more difficult than those with an unknown compare/result/whole set (Gabler & Ufer, 2020; Van Lieshout & Xenidou-Dervou, 2020). This may be connected to the mathematical structure, which is determined by the unknown set: For the latter type of unknown set, learners can construct a mathematical model with an operation, which is directly applicable (e.g., 7 + 8 = x), while the operation resulting from the other type of unknown set may be represented implicitly in learners' mathematical models (e.g., x + 8 = 15). Learners may subsequently transform this implicit representation into a directly applicable mathematical structure (e.g., 15–8 = x).

2.1.4. Unknown set and a/s wording

The influence of the unknown set on a word problem's difficulty is also connected to the a/s wording (Briars & Larkin, 1984). Word problems in which the directly applicable mathematical operation is inconsistent with the wording are usually harder than consistent word problems ("consistency hypothesis", Lewis & Mayer, 1987). For example, the inconsistent word problem "Susi has three marbles. She has two marbles less than Max. How many marbles does Max have?" contains a *subtractive* wording ("less"), but *addition* is directly applicable (3 + 2 = x). The findings on the consistency of word problems are supported by eye tracking studies, which observed that some learners mainly focus on key words and deduce the mathematical operation directly from this operation (Hegarty et al., 1992). Solving inconsistent word problems requires a deep analysis of the situation and therefore the construction of a sound situation model (Scheibling-Sève et al., 2020). Although most studies on the consistency of word problems focused on early primary school children, Verschaffel (1994) could still confirm the consistency hypothesis with 10–11 year olds. This underlines the importance of finding effective instructional approaches to support learners with the understanding of the given situation structure already in early school years.

2.2. Language skills as prerequisite of word problem solving

Beyond students' domain-general abilities (e.g., fluid intelligence), their subject-specific skills and knowledge, their affect (e.g., the emotion of feeling motivated or frustrated), and their social background, language skills have been discussed as a prerequisite of word problem solving (Verschaffel et al., 2015; Vilenius-Tuohimaa et al., 2008).

Language skills have been shown to influence learning mathematics, since learners need these skills to construct mathematical knowledge and to participate in classroom discourse (Moschkovich, 2015). In the context of word problem solving, language skills are particularly required in the form of reading skills since arithmetic situations are represented in written form. This is also confirmed by the meta-analysis by Peng et al. (2020), who could observe a stronger relation between language and mathematics skills for word problem solving in comparison to other mathematical skills. Lower-level technical decoding skills (e.g., reading accuracy, reading fluency) as well as higher-level reading comprehension skills (in the sense of textual understanding) are crucial to decode the text base and to construct an accurate situation model (Vilenius-Tuohimaa et al., 2008). Studies in primary and secondary schools have repeatedly identified technical decoding skills and reading comprehension skills to be significant predictors of performance in word problem solving (Muth, 1984; Vilenius-Tuohimaa et al., 2008). Students with high reading comprehension skills seem to create more accurate situation models and mathematical models (Leiss et al., 2010).

2.3. Flexibility in dealing with additive situations

When solving a word problem, learners initially decode the words and sentences of the text base and integrate this information into a situation model (Kintsch & Greeno, 1985). In the best case, this initial situation model contains all features of the situation structure, which were realized in the text base. The text base can emphasize different *features of the situation structure*. For example, the fact that Max has four objects (e.g., marbles, cookies) more than Susi can also be described by saying that Susi would have as many objects as Max, if she received four more objects from someone else. An original comparison of sets would be interpreted as an equalization in this new perspective.

According to models of reading comprehension (e.g., Kintsch, 1998), learners do not only reconstruct the content of the text base, but may also add inferences to their situation model by restructuring their situation model with features of the situation structure. For example, learners could *restructure* a compare situation by inferring a dynamic perspective, in which the difference between the two compared sets is equalized by an action, and integrate this new perspective into their individual situation model.

Selecting a particular mathematical structure as a mathematical model is based on the learners' situation model. However, identifying a mathematical structure also depends on the learners' conceptual knowledge: Learners need conceptual knowledge about how different features of a situation structure relate to certain mathematical concepts. Conceptual knowledge, in this sense, comprises "principles that govern a domain and the interrelations between units of knowledge in this domain" and is expected to help students to organize "information in their internal representations of [the] problems" (Rittle-Johnson et al., 2001). Psychological models use the term "schemata" to describe the mental representation of situation structures (as situation models) and solution strategies (as mathematical models) (Kintsch & Greeno, 1985). Common to both perspectives is the assumption that different "schemata" must be individually available and also activated, so that learners can mathematize situation models. For instance, in a study with 21 German 9th grade classes, Leiss et al. (2010) identify the construction of an adequate situation model as a specific difficulty generating feature in word problem solving. Depending on which features of the situation structure are available in the students' individual situation model, different schemata can be activated to generate a mathematical model.

In our context, even if learners can represent a compare situation structure (Riley et al., 1983; Schipper, 2009) adequately in their individual situation model, inferring additional features of the situation structure (e.g., inferring a dynamic perspective) may allow them to identify the correct mathematical model even with restricted conceptual knowledge how compare situations relate to addition and subtraction.

Riley et al.'s (1983) seminal (computer) model of word problem solving, for example, proposed that compare situations could be restructured as combine situations, by establishing a one-to-one correspondence between the reference set and an equally large subset of the comparison set (or vice versa), and identifying the difference set as the other subset. Indeed, classical results by Hudson (1980) showed that, if compare problems are re-worded so that this one-to-one correspondence occurs explicitly, the word problem's difficulty reduces substantially for children up to grade 1. Even though potentially very useful, restructuring a compare situation as a combine situation requires a substantially new perspective on the situation that is often hard to establish within the described context in a meaningful way. Depending on which information is available in the word problem and the initial situation model, other restructurings might be more helpful. The perspective proposed in this manuscript is to focus not on one potentially "most useful" restructuring strategy for each problem type, but to provide learners with a network of connections between different problem types that allows for flexible (and in the best case adaptive) restructuring of word problems in different ways.

Thus, choosing an adequate mathematical operation is contingent on which features of a situation structure (cf. Chapter 2.1.) are included in the students' situation models: Based on the features presented in the problem text alone, it may be more or less straightforward to construct a mathematical model. To overcome the reported barriers to mathematize their individual situation model, learners may benefit from inferring further features to the situation model based on their own knowledge about additive situations. We assume that this, in turn, may facilitate choosing an adequate mathematical operation. Based on the idea that conceptual knowledge comprises knowledge about interrelations (Rittle-Johnson et al., 2001), it may be helpful for learners to expand their conceptual knowledge by analyzing connections between different additive situation structures.

In summary, constructing a situation model plays a central role in word problem solving. This can be described by theoretical models such as the text comprehension model by Kintsch (1998), which emphasizes the importance of reducing, organizing, and elaborating information from the text base, and it is also supported by empirical evidence (e.g., Leiss et al., 2010; Stern & Lehrndorfer, 1992; Thevenot et al., 2007). This work assumes that it is not only the decoding of the text base that is responsible for these differences in difficulty ("comprehension obstacles", Prediger & Krägeloh, 2015), but primarily the identification of a mathematical structure that matches the identified situation structure ("conceptual obstacles").

In order to facilitate the latter process of overcoming conceptual obstacles, research from the early eighties suggested introducing strategies to restructure situation models with further information, so that they can be mathematized more easily (Fuson et al., 1996; Greeno, 1980; Stern, 1993). In this paper, strategies are understood as cognitive procedures that have a heuristic value when solving a certain type of problem. In the following, we are going to elaborate on two strategies (beyond the one from the Riley et al., 1983 model) that focus on restructuring the situation model by adding inferences: The *Inversion Strategy*, which is primarily targeted to turn inconsistent word problems into consistent ones, and the *Dynamization Strategy*, which is primarily targeted to turning difficult, static compare problems into potentially easier, dynamic equalize problems (Fig. 2).

2.3.1. Inversion Strategy: changing the perspective on mathematical relations

Stern (1993) and other researchers (Fuson et al., 1996; Verschaffel, 1994) stress the role of understanding relational statements and the perspective on these relations. Stern (1993) found that 70% of the interviewed first graders did not identify relational statements such as "Max has 5 marbles more than Susi" and "Susi has 5 marbles less than Max" as equivalent. However, understanding this linguistic *symmetry* of relations may support students in solving compare problems (Stern, 1993). Flexibly switching between linguistically symmetrical statements (*inverting* the direction of the relational term) may allow students to restructure more difficult compare problems with an unknown reference set as empirically easier ones with an unknown compare set and thus transform inconsistent into consistent word problems (see Chapter 2.1.4., "Consistency Hypothesis", Lewis & Mayer, 1987). To apply the Inversion Strategy, students need to reverse subject and object, and invert the a/s wording.

2.3.2. Dynamization Strategy: changing the semantic structure

Another suggestion aims at restructuring difficult semantic structures as easier accessible structures. Greeno (1980) proposed to help learners with restructuring the semantic structure of change problems such as "Jill had 3 apples. Betty gave her some more apples. Now Jill has 8 apples. How many did Betty give her?" as a combine situation with "3" as part and "8" as whole. Considering the reported difficulties of students when solving compare problems, this idea could be transferred to a similar strategy. Restructuring static compare problems into dynamic equalize problems seems particularly obvious here, since these situations both contain disjoint sets, but equalize problems are considered easier than compare problems in models of word problem difficulty (Nesher et al., 1982). Also Fuson et al. (1996) reported higher solution rates for equalize problems in comparison to compare problems in a study with first and second graders. Thus, students may *dynamize* compare problems by restructuring them as equalize problems, making them easier to represent and mathematize.

Defining the construct. These two strategies may be one way to restructure the learner's situation model by adding inferences. We define this *flexibility in dealing with additive situations* as the skill to restructure situation models of additive one-step word problems by inferring additional features of the situation structure (a/s wording, semantic structure, unknown set). This includes restructuring a described situation regarding its situation structure, inferring features of the situation structure that are not described in the text base, and deciding if a description fits the verbally presented situation or not. If learners struggle during word problem solving, they may use this skill to spontaneously restructure their knowledge (as described in the theory of "cognitive flexibility", Spiro et al., 1991).

Learners with high flexibility may generate descriptions of the given situations that (1) describe the situation accurately, and (2) infer additional features of the situation structure (e.g., an alternative description of the a/s wording). Such restructured situation models may support learners with constructing mathematical models. A low flexibility in dealing with additive situations may be one reason for difficulties with certain word problem types. If the described flexibility could be fostered, this may also help students when encountering difficult word problem types. Practical observations (Fromme et al., 2011) as well as theoretical considerations (Greeno, 1980; Stern, 1993) endorse approaches based on advancing learners' flexibility. The approach differs from established approaches, which primarily offer one standard way to interpret a specific type of word problems, e.g., as a part-whole situation (Riley et al., 1983; Wolters, 1983). Dynamization and Inversion Strategies both aim at restructuring the learners' situation model with further inferences

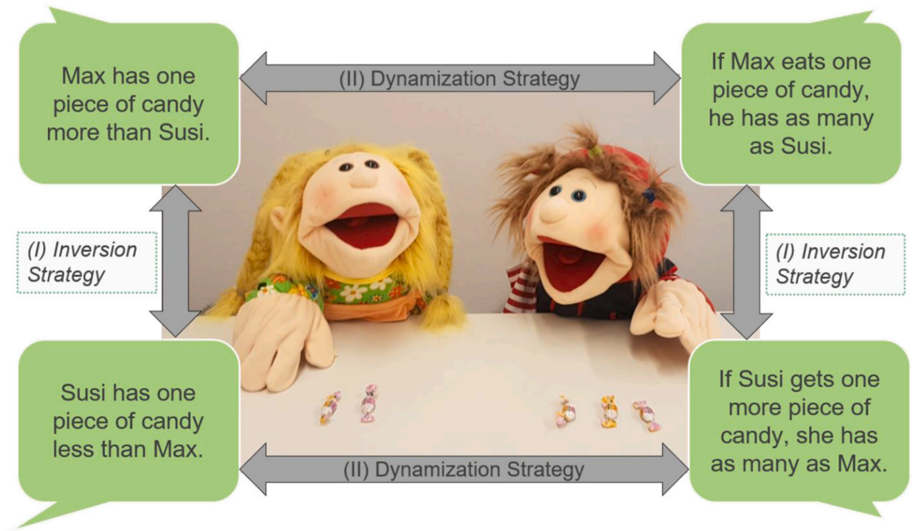

Fig. 2. Examples for inversion strategy and dynamization strategy.

while reconstructing the situation structure, so that the situation model can be mathematized more easily. The two strategies may complement conceptual knowledge, which is necessary to solve word problems (Scheibling-Sève et al., 2020). They may help learners to focus on relevant features of the situation structure and to infer further features (Rittle-Johnson et al., 2001). Conveying these strategies may be one way to develop flexibility.

3. The current study

The current study aims to investigate, if conveying strategies to restructure situation models, such as the Inversion and the Dynamization strategy, is a feasible way to foster students' flexibility in dealing with additive situations and word problem solving skills. While offering standard ways to interpret specific problem types has been effective (cf. Verschaffel et al., 2007), we are not aware of studies investigating trainings, which provide students with a range of strategies to restructure situation models and explicitly investigate the resulting flexibility in dealing with additive situations. Thus, we were interested, if such a training would benefit students *at all*. However, we were careful not to include any tasks in the training that could practice other aspects of word problem solving beyond dealing with concrete situations (i.e. generating mathematical models, calculating or interpreting results). We addressed the following questions:

RQ1: Does training strategies to restructure situation models go along with progress in students' flexibility in dealing with additive situations (RQ1a) and their word problem solving skills (RQ1b) positively? Are these effects still present after four weeks?

The strategies conveyed in the training have been proposed as potentially helpful for restructuring situation models. Thus, we expected training them would benefit students' flexibility. As suggested by Stern (1993) and Greeno (1980), we expected that students would also make progress based on these strategies when solving word problems. Only few word problem training studies investigate (and find) sustained effects (Lein et al., 2020). Thus, we did not pose hypotheses regarding long-term effects.

RQ2: Is students' different progress regarding flexibility in dealing with additive situations (RQ2a) and word problem solving skills (RQ2b) in the experimental and control condition related to students' language skills directly after the training, and after four weeks?

Learners with low language skills may face specific problems when learning to solve word problems (Peng et al., 2020). Our training aims to restructure situation models by re-wording word problems. Specific care was taken to support learners with low language skills regarding the linguistic challenges of this approach. Consequently, the training may particularly help learners to overcome the specific challenges posed by low language skills. Learners with high language skills may already have other strategies to approach word problems, to which the conveyed strategies are not well aligned, possibly resulting in a so-called "expertise reversal effect" (Kalyuga et al., 2012). Additional cognitive resources might be needed to match their existing strategies with those from the training (Kalyuga et al., 2012). These considerations speak for a stronger effect for learners with low language skills than for learners with high language skills. However, if the measures to mitigate the specific linguistic challenges of the conveyed strategies are not effective, high language skills may be required to enact the conveyed strategies and use them to restructure situation models. This would result in higher benefits for learners with high language skills, similar to a Matthew effect (Merton, 1968). Finally, the training might also be similarly effective for all students, regardless of their language skills. Prediger and Wessel (2018) reported comparable effects for learners with low and high language skills in a training on fractions.

RQ3: Do the differences regarding the development of students' word problem solving skills between control and experimental group remain, when statistically controlling for students' concurrent flexibility in dealing with additive situations?

We assumed that the training's effects on word problem solving would be caused due to students' improved flexibility. Thus, we expected that the effects would deteriorate, if students' flexibility, measured at the same time as the corresponding word problem score, was controlled statistically.

4. Method

4.1. Procedure and design

We conducted an experimental training study with second graders. According to the Bavarian curriculum, second graders have already practiced to connect mathematical operations with actions and situations. They are able to use this information to explain the connection between addition and subtraction (e.g., increasing a set and then reversing this action). The curriculum does not mention compare problems despite their empirical difficulty. Also textbook analyses (e.g., Gabler et al., 2023) indicate that compare problems occur barely in math lessons. School authorities approved the study, guardians gave consent to their child's participation in the study. Around mid-term, students answered a pretest on their general cognitive abilities, reading comprehension skills, basic arithmetic skills, flexibility in dealing with additive situations, word problem solving skills, and personal data. Based on the pretest, students were randomly allocated to experimental and control group. The experimental group received training, while the control group participated in regular instruction (subjects different from mathematics, mostly arts and German courses). The posttest directly after the training and the follow-up four weeks later surveyed students' flexibility in dealing with additive situations and word problem solving skills (Fig. 3).

4.2. Sample and group assignment

The initial sample comprised $N = 130$ s graders from ten classrooms in three primary schools located in Munich, Germany, who attended the first measurement. Students were randomly assigned to the experimental or control group, considering students' reading comprehension scores. Within each classroom, students' reading comprehension scores were ranked. Out of the six highest scores, we randomly picked three students. The same procedure was repeated for the six lowest scores, resulting in six students per classroom forming a training group. All other remaining students of the classroom were assigned to the control group. With this procedure, we aimed to generate training groups with heterogeneous language skills, and to ensure comparability with the control group. Furthermore, learners with low language skills may benefit from contributions made by learners with high language skills (Pyle et al., 2017). This resulted in a control group ($N = 70$) and ten training groups (one per classroom) of six students each ($N = 60$).

Those students taking part at the pretest, posttest, and follow-up were included in the formal analyses ($N = 115$, $N = 54$ experimental group; $N = 61$ control group). A sensitivity power analysis (repeated measures ANOVA, within-between interaction) for $\alpha = 0.05$ and $\beta = 0.95$ yielded an effect size of $f = 0.15$ (small to medium effect).

4.3. Training

4.3.1. Organization

The training comprised ten 40–50 min small-group sessions over five weeks. Pre-service teachers acted as tutors, following a script outlining content, procedure, and duration of the activities as well as wording options for student support. The training was piloted with $N = 4$ students from another school.

4.3.2. Content

The training aimed at developing flexibility in dealing with additive situations. No word problems were solved during the training, which means that students did not encounter word problems in their typical form (e.g., as in the word problem test) and were not asked to write down equations or make calculations. The program contained learning activities regarding difference sets (using the restructuring of compare situations as combine situations; Riley et al., 1983), the Dynamization Strategy and the Inversion Strategy. After an initial phase of familiarizing with certain basics (e.g., relational statements), the Dynamization Strategy and the Inversion Strategy were introduced by dealing with *given* statements based on Fig. 2. One activity type involved verifying, if a statement corresponds to one given situation, another matching a statement with several given situations. In the second half of the training, learners were encouraged to generate different descriptions of given situations. Fig. 4 gives an overview of the program's structure. More detailed information on the training program as well as case-studies on its implementation and uptake by students is provided in (Gabler & Ufer, 2021).

During the program, learners were encouraged to contrast and analyze different features of given statements collaboratively. We intended to provide learners with a cognitive tool to restructure their situation models with further features, when solving word problems later. Learners were also asked to explain differences and commonalities between different statements (e.g., two symmetrical relational statements). Engaging in activities such as explaining, arguing, and justifying statements or descriptions was intended to enhance rich discourse practices (Erath et al., 2021).

4.4. Instruments

Trained university student assistants administered all instruments following explicit guidelines. Students completed all instruments as paper-pencil tests in individual work. Each test instrument was introduced by sample tasks. After phases of 15–25 min, short relaxation

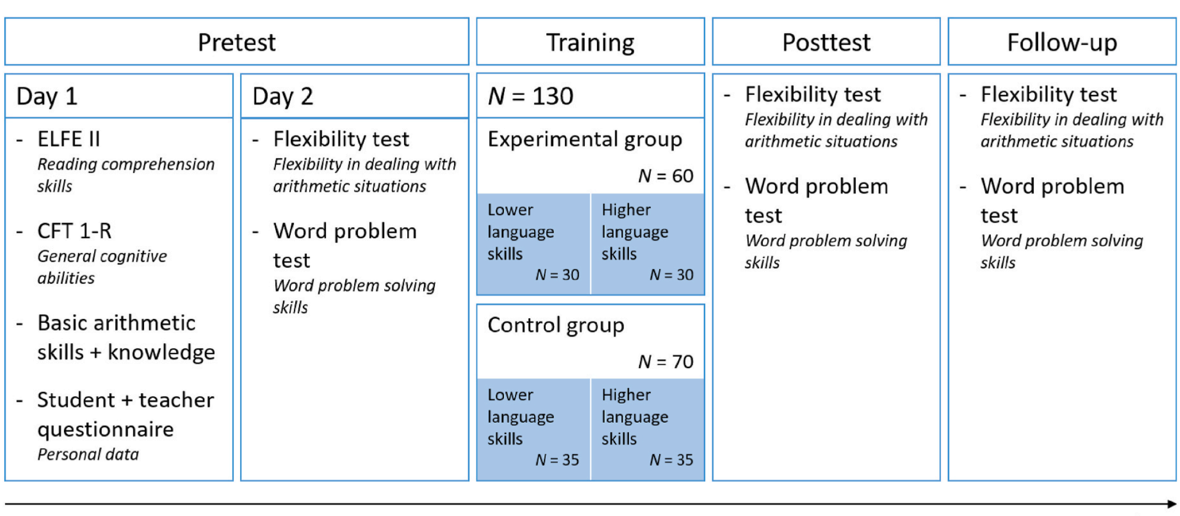

Fig. 3. Overview of the training study.

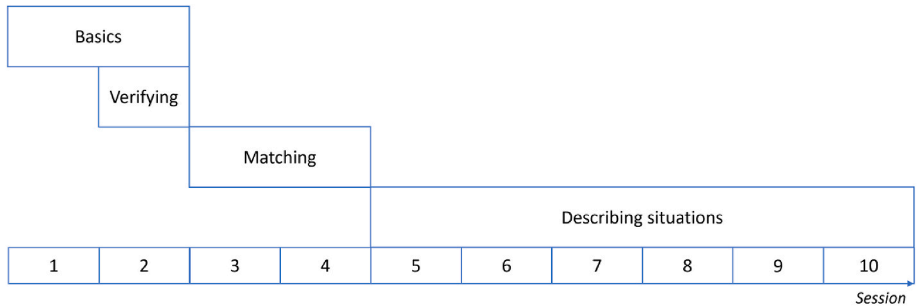

Fig. 4. The training program's structure.

breaks with physical activities were included. The supplementary materials provide additional details on the self-developed instruments.

4.4.1. *Flexibility in dealing with additive situations*

To measure flexibility in dealing with additive situations, a new instrument was developed, pilot-tested (Gabler & Ufer, 2022), and adapted afterwards. The items were embedded into a story about two twins, who report on a birthday party they visited (Fig. 5).

The learners decided, if the twins' statements say the same thing. The statements emphasized different features of the situation structure. Some statement pairs referred to different situations (Fig. 5). Statements pairs that were equivalent corresponded either with the Dynamization Strategy, contrasting different semantic structures, or the Inversion Strategy, contrasting different a/s wordings. The test comprised eight Dynamization items and twelve Inversion items (testing time: 15 min). The answers were scored dichotomously. Missing and "I don't know" (pretest: 4.5%, posttest: 3.0%, follow-up: 2.5%) answers were treated as incorrect. The reliability was satisfying (pretest: $\alpha = 0.80$, posttest: $\alpha = 0.82$, follow-up: $\alpha = 0.88$.

4.4.2. *Word problem solving skills*

The word problem test was adapted from prior studies by Stern (1998). Systematically varying the semantic structure, a/s wording, and unknown set resulted in 20 word problem types in total (Gabler & Ufer, 2020), represented by one item each (see Fig. 6 for a sample item).

Each student worked on one of six booklets each containing a subset of 12 out of the 20 items (testing time: 12 min). The different booklets covered all 20 items and were linked in a multi-matrix design. Students' solutions were scored as correct, if the numerical result was correct. Since not all students answered all items, scores on a common scale were acquired by scaling data with a one-dimensional Rasch Model (Rasch, 1960). This model estimates a performance parameter on a latent scale for each learner and can provide comparable performance parameters also in the situation of a multi-matrix design, where participants worked different item sets. Mean person scores were restrained to zero across all three measurements. The WLE reliability was 0.63.

4.4.3. *Language skills*

The students' language skills were measured with the German reading comprehension test ELFE II (Lenhard & Schneider, 2018). It assesses receptive language skills based on reading fluency and accuracy in a group setting. As recommended for second graders in the test

Fig. 5. Translated sample item of the flexibility test.

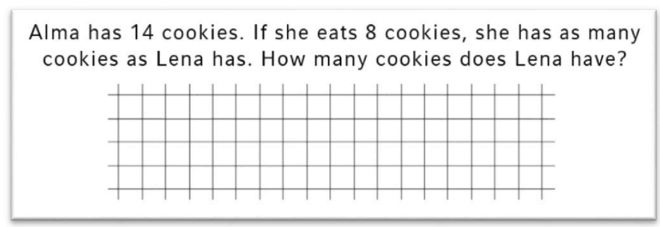

Fig. 6. Sample item of the word problem test.

manual, we used the subscales on the word level (choosing one out of four words which matches the given picture; speed test, 3 min) and sentence level (choosing one out of four words that fits into the given sentence; speed test, 3 min). Answers were coded dichotomously. The reliability was excellent ($\alpha = 0.97$).

4.4.4. *General cognitive skills*

General cognitive skills were measured by using the subscales "Similarities", "Classifications", and "Matrices" of the Culture Fair Intelligence Test (CFT 1-R, Weiβ & Osterland, 2013), which measure characteristics of general cognitive skills in a culturally fair, language-free setting. The reliability of the three subscales was acceptable (subscale "Similarities": $\alpha = 0.66$; "Classifications": $\alpha = 0.73$; "Matrices": $\alpha = 0.80$). The three subscales were combined into one joint indicator.

4.4.5. *Basic arithmetic skills*

Basic arithmetic skills were measured with a test by Bochnik (2017), that was adapted for second graders in this study. Some of the tasks relate to technical skills in adding and subtracting numbers ranging until 100. Further tasks required conceptual knowledge, for example on the relationship between addition and subtraction (e.g., by asking for all four calculations that can be conducted with the numbers 7, 8, and 15). The reliability is satisfying ($\alpha = 0.82$).

4.5. *Statistical analyses*

Repeated measures ANOVAs were conducted based on linear mixed models with Bonferroni-corrected post-hoc tests. Dependencies in the clustered sample (students within classrooms) were modeled as a random factor. The calculations were executed in R with the packages *lme4*, *emmeans*, *effectsize,* and *insight.*

5. Results

5.1. *Preliminary analyses*

See Table 1 for descriptive statistics of all measures. ANOVAs did not indicate significant treatment group differences at pretest for general cognitive skills, reading comprehension, basic arithmetic skills, initial

Table 1
Descriptive statistics for all measures.

Measure	*N*	*M*	*SD*	*skew*	*kurtosis*	*min*	*max*
General cognitive skills	115	0.67	0.13	−0.55	0.99	0.22	0.98
Reading comprehension	115	45.12	15.52	0.20	−0.31	6.00	87.00
Basic arithmetic skills	115	7.46	3.84	0.14	−0.90	0.00	16.00
Flexibility T1	115	14.06	4.06	−0.53	−0.51	3.00	20.00
Flexibility T2	115	15.75	3.86	−0.90	−0.10	4.00	20.00
Flexibility T3	115	16.15	4.17	−1.49	2.00	0.00	20.00
Word Problems T1	115	−0.27	1.73	0.23	−0.12	−4.36	4.32
Word Problems T2	115	−0.19	1.70	0.13	0.02	−4.30	4.32
Word Problems T3	115	0.13	1.72	−0.03	0.92	−5.00	4.32

Note. *N*: number of participants, *M*: mean, *SD*: standard deviation, *min*: minimum, *max*: maximum.

flexibility, or initial word problem solving skills ($F(1, 113) < 0.23$, $p > 0.63$). General cognitive skills and basic arithmetic skills were not included in covariates in the subsequent analyses. Including them did not change the overall pattern of effects.

5.2. *Flexibility in dealing with additive situations (RQ1a, RQ2a)*

Fig. 7 and Table 2 show the descriptive results in the flexibility test for students from the experimental and the control group, differentiated by low vs. high language skills.

A repeated measures ANOVA (Table 3) showed a significant interaction effect of *measurement time* and *treatment group* with a small effect size ($F(222.00, 2) = 3.17$, $p = 0.044$, $\eta_p^2 = 0.03$). This indicates that the experimental group and control group developed differently along the measurements (RQ1a). The other interaction effects, including the three-way interaction of *language group*, *measurement time*, and *treatment group*, were not significant, indicating that the treatment effect on students' flexibility development was not statistically different for students with low vs. high language skills (RQ2a).

5.3. *Word problem solving skills (RQ1b, RQ2b)*

Fig. 8 and Table 4 show the descriptive results in the word problem test for students from the experimental and the control group, differentiated by low vs. high language skills.

A repeated measures ANOVA (Table 5) showed a significant interaction effect of *measurement time* and *treatment group* with a small effect size ($F(222.00, 2) = 3.45$, $p = 0.034$, $\eta_p^2 = 0.03$). This indicates that the experimental group and control group developed differently along the measurements (RQ1b). The other interaction effects, including the three-way interaction of *language group*, *measurement time*, and *treatment group*, were not significant, indicating that the treatment effect on students' flexibility development was not statistically different for students with low vs. high language skills (RQ2b).

The experimental group showed significant performance growth from pretest to follow-up, but not from pretest to posttest, while the control group showed no significant progress (*experimental:* pre–post: $b = 0.26$, $p = 0.93$, $d = 0.16$; pre–follow-up: $b = 0.81$, $p < 0.001$, $d = 0.52$; *control:* pre–post: $b = -0.08$, $p = 1$, $d = -0.05$; pre–follow-up: $b = 0.04$, $p = 1$, $d = 0.03$). The performance growth between pretest and follow-up was significantly stronger in the experimental than in the control group

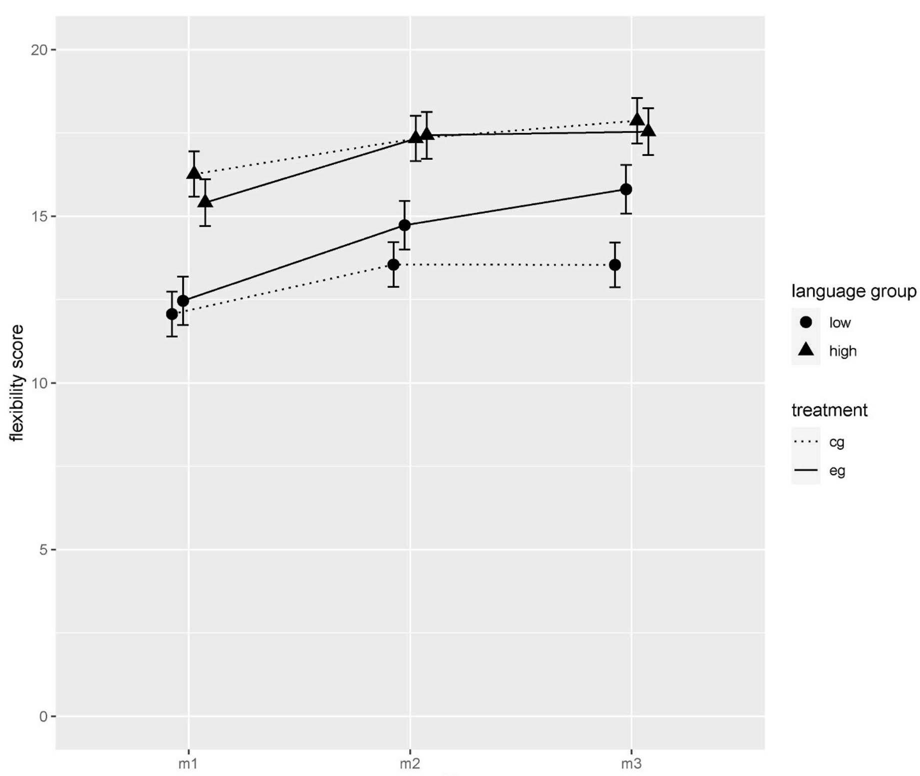

Fig. 7. Development of flexibility scores by treatment and language group (estimated marginal means and standard errors).

Table 2
Descriptive results for the flexibility test by treatment group, language group, and measurement.

Treatment	Language	*N*	T1		T2		T3	
			M	*SD*	*M*	*SD*	*M*	*SD*
Experimental	high	28	15.41	3.36	17.43	2.27	17.54	2.38
Experimental	low	26	12.46	3.54	14.73	3.54	15.81	3.54
Control	high	30	16.27	2.92	17.33	3.17	17.87	3.17
Control	low	31	12.06	4.57	13.55	4.56	13.54	5.39

Note. *N*: number of participants per group, *M*: mean, *SD*: standard deviation.

Table 3
Repeated measures ANOVA for the flexibility test.

Factor	*df1*	*df2*	*F*	*p*	η_p^2
Measurement	2	222.00	42.21	<0.001***	0.28
Treatment	1	110.46	0.54	0.463	0.00
Language	1	110.70	27.55	<0.001***	0.20
Measurement x treatment	2	222.00	3.17	0.044*	0.03
Measurement x language	2	222.00	0.63	0.536	0.01
language x treatment	1	105.74	1.73	0.191	0.02
Measurement x treatment x language	2	222.00	1.41	0.246	0.01

Note. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Experimental and control group showed significant flexibility growth from pretest to posttest resp. follow-up (*experimental:* pre–post: $b = 2.15$, $p < 0.001$, $d = 0.58$; pre–follow-up: $b = 2.74$, $p < 0.001$, $d = 0.75$; *control:* pre–post: $b = 1.28$, $p < 0.001$, $d = 0.35$; pre–follow-up: $b = 1.54$, $p < 0.001$, $d = 0.42$). The performance growth between pretest and follow-up was significantly stronger in the experimental than in the control group ($b = 1.20$, $p = 0.03$, $d = 0.33$), but not from pretest to posttest ($b = 0.87$, $p = 0.16$, $d = 0.24$).

($b = 0.77$, $p = 0.02$, $d = 0.49$), but not from pretest to posttest ($b = 0.34$, $p = 0.50$, $d = 0.21$).

5.3.1. Exploratory analysis per problem type

Due to the restricted power of our study and the low number of items, we investigated students' progress per problem type only descriptively in an exploratory manner (Fig. 9), so interpretations should be made with care. The two combine items were excluded, since they were very easy for all participants. Indeed, we observed the descriptively strongest increase in the experimental group for compare problems. From m1 to m3, is increase is descriptively stronger than for the other two problem types in the experimental group, and also stronger than for compare problems in the control group.

5.4. Effects on word problem solving skills after controlling for flexibility (RQ3)

To investigate RQ3, students' flexibility in dealing with additive situations (at the respective measurements) was added as an additional covariate in the analyses in 4.3 (Table 6).

Flexibility explained a significant share of variance in word problem solving scores ($F(206.39, 1) = 62.28$, $p < 0.001$, $\eta_p^2 = 0.23$). Neither the effect of *measurement time* ($F(230.33, 2) = 1.76$, $p = 0.1754$, $\eta_p^2 = 0.02$)

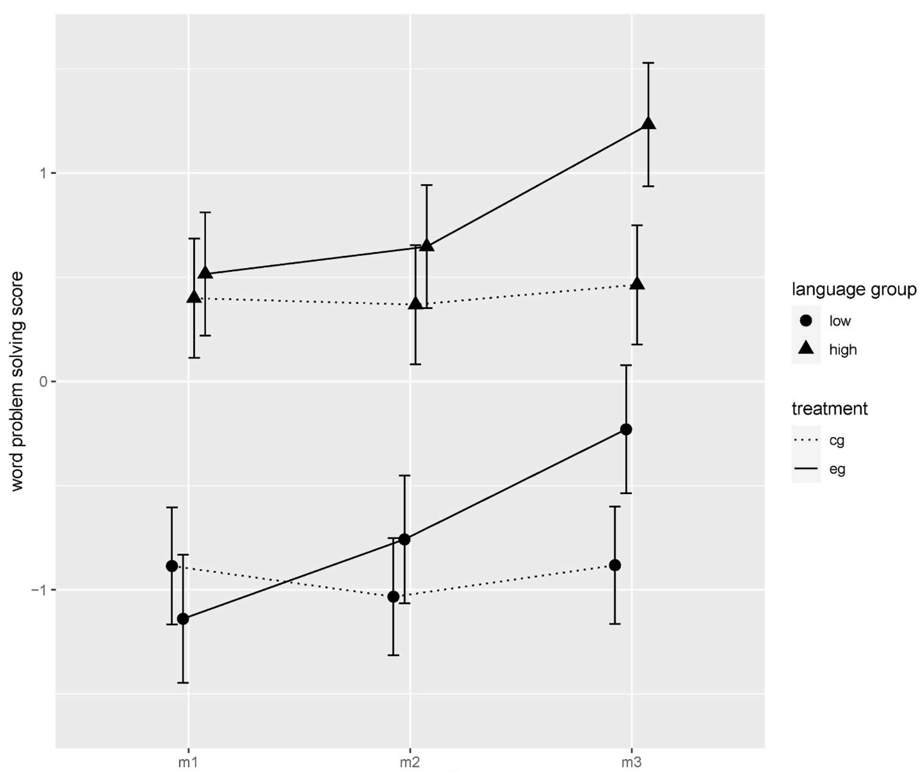

Fig. 8. Word problem solving scores for experimental and control group by language skills for the three measurements (estimated marginal means and standard errors).

Table 4
Descriptive results for the word problem test by treatment group, language group, and measurement.

Treatment	Language	N	T1		T2		T3	
			M	*SD*	*M*	*SD*	*M*	*SD*
Experimental	high	28	0.52	1.42	0.65	1.39	1.23	1.93
Experimental	low	26	−1.14	1.64	−0.76	1.49	−0.23	1.50
Control	high	30	0.40	1.73	0.37	1.66	0.46	1.15
Control	low	31	−0.90	1.52	−1.03	1.68	−0.88	1.53

Note. *N*: number of participants per group, *M*: mean, *SD*: standard deviation.

Table 5
Repeated measures ANOVA for the word problem test.

Factor	*df1*	*df2*	*F*	*p*	η_p^2
Measurement	2	222	4.69	0.010***	0.04
Treatment	1	111	1.68	0.198	0.01
Language	1	111	36.07	<0.001***	0.25
Measurement x treatment	2	222	3.45	0.034*	0.03
Measurement x language	2	222	0.04	0.960	0.00
language x treatment	1	111	0.11	0.739	0.00
Measurement x treatment x language	2	222	0.19	0.830	0.00

Note. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

nor its interaction with the *treatment group* ($F(215.52, 2) = 1.73$, $p = 0.180$, $\eta_p^2 = 0.02$) were significant anymore. Removing the three-way interaction as well as the two-way interaction between measurement and treatment from this model did not significantly affect overall model fit ($\chi^2(4) = 4.19$, p = 0.38).

6. Discussion

We proposed to foster students' flexibility in dealing with additive situations to develop their word problem solving skills and investigated this approach in a feasibility study. We will discuss the following main findings: It is possible to develop students' flexibility in dealing with additive situations (6.1). The training also has effects on students' word problem solving skills (6.2). The effects occurred, however, only at follow-up (6.3). They are not significantly different for learners with high vs. low language skills (6.4). The training's effects on students' word problem solving skills can be partially explained by its effects on students' flexibility (6.5).

6.1. *Flexibility in dealing with additive situations (RQ1a)*

Strategies to restructure situation models have been proposed (Fuson et al., 1996; Greeno, 1980; Stern, 1993), but empirical evidence was scarce. In our study, students showed significant progress regarding flexibility in dealing with additive situations. Even though this may partially be due to increasing familiarity with the flexibility test, it indicates that repeatedly analyzing additive situations may relate to a positive development. Specifically, training strategies to achieve this flexibility proved effective in our study compared to a neutral control. Prior qualitative results showed that some students struggle substantially with the training content, due to low language skills and missing conceptual knowledge for example regarding difference sets (Gabler & Ufer, 2021). The study shows that, on average, a positive development can be achieved by targeted training. It remains open, to which extent personalized support may be necessary for students with less beneficial prerequisites (Dumont & Ready, 2023).

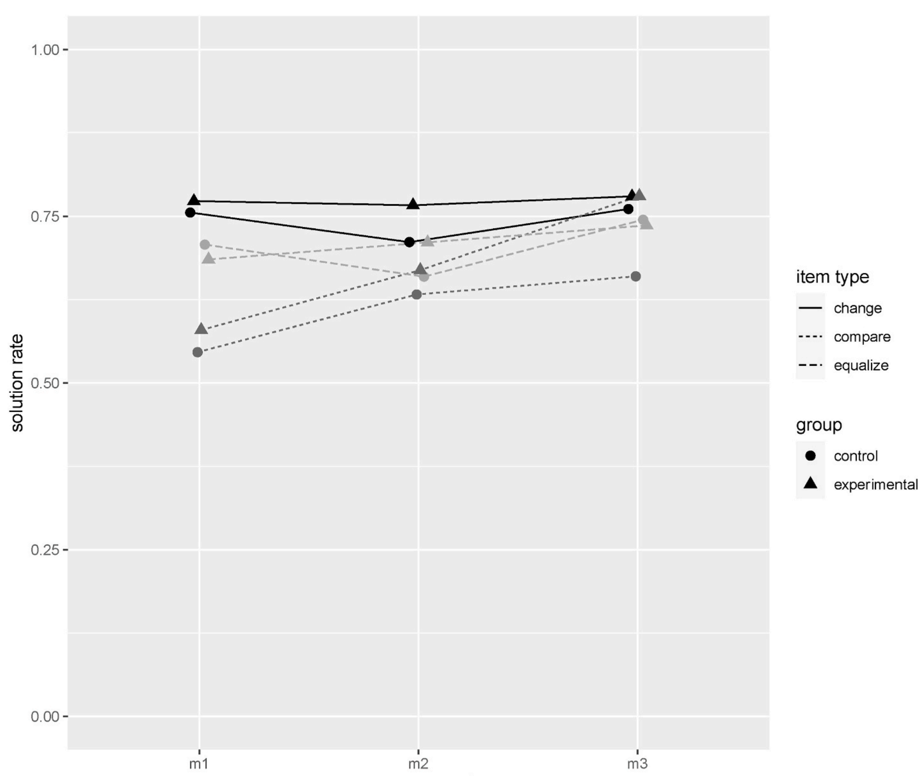

Fig. 9. Word problem solving solution rates for experimental and control group by problem type for the three measurements.

Table 6
Repeated measures ANOVA for the word problem test, with flexibility as a covariate.

Factor/covariate	*df1*	*df2*	*F*	*p*	η_p^2
Measurement	2	230.33	1.76	0.175	0.02
Treatment	1	103.23	1.46	0.230	0.01
Language	1	113.45	17.02	<0.001***	0.13
Measurement x treatment	2	215.52	1.73	0.180	0.02
Measurement x language	2	214.40	0.01	0.991	0.00
language x treatment	1	103.73	1.53	0.219	0.01
Measurement x treatment x language	2	214.75	0.28	0.760	0.00
flexibility	1	206.39	62.28	<0.001***	0.23

Note. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

6.2. *Word problem solving skills (RQ1b)*

Also the performance of the experimental group regarding word problem solving skills increased stronger than in the control group. This adds a new approach to existing trainings on word problem solving (Lein et al., 2020; Verschaffel et al., 2020). The neutral control group in our study only allows to conclude that fostering flexibility is a *feasible* way to develop word problem solving. Based on this study, we cannot draw conclusions on its effectiveness compared to other training approaches such as schema-based instruction (Cook et al., 2020).

6.3. *Primarily delayed effects (RQ1)*

Training effects on both outcomes were not significant directly after the training. This is of particular interest, since sustained effects are rarely investigated or found in research on word problem solving (Lein et al., 2020). One reason, why the descriptively smaller immediate effects of the training did not reach significance, could be the restricted power of our study. Future research should go beyond small feasibility studies here. However, our study indicates that the effects of the training are not restricted to the time when it is administered. We can exclude that further learning opportunities on word problem solving caused this growth: All mathematics teachers reported that they focused on introducing multiplication and axial symmetry in geometry, and not on additive word problems. Second, due to the randomized group assignment, it is plausible that the effects can be traced back causally to the training. Delayed progress after the training could also be due to application and elaboration of the acquired, but not yet sufficiently routinized strategies during everyday life. Students do focus on numbers and relations spontaneously also outside of school (spontaneous focusing tendencies, McMullen et al., 2013), which may have given them the opportunity to extend their newly gained skills, for example dynamizing a static situation or inverting the a/s wording of relational statements, even after the training ended. If this explanation can be substantiated, it would be a convincing sign towards the effectiveness of the trained strategies.

6.4. *Role of language (RQ2)*

Prediger and Wessel (2018) pointed out that the effectiveness of mathematics and language trainings is commonly investigated for learners with low language skills, while findings on learners with high language skills are scarce. We investigated matched pairs of learners with low *and* high language skills. Different arguments spoke for different patterns of influence of language skills on the training's effect. The results showed no indication that the training effects depended on language skills. Based on the sensitivity power analysis, this indicates that the benefit for learners with high and low language skills differs at most by a small to medium effect. This adds to findings by Prediger and Wessel (2018), who also could not identify differential instructional needs of learners with different language skills. It is not clear yet, which mechanisms can explain this effect. It is possible that two opposite mechanisms are in balance: On the one hand, the training activities may indeed address skills that learners with low language skills struggle most with (Peng et al., 2020), providing a potential advantage for them. At the same time, learners with high language skills may be able to use the entailed learning opportunities better (Merton, 1968). The exact mechanisms need to be investigated in future research. It is, however, promising that learners could benefit from the training regardless of their language skills.

6.5. *Flexibility explains gains in word problem solving skills (RQ3)*

Finally, we explored if the gains in flexibility could explain parallel gains in word problem solving skills. Indeed, the effects of temporal development and of the training were rendered insignificant, when students' flexibility was considered. We take this as first (though, weak) evidence that flexibility development explains interindividual differences in word problem solving skills. This would be in line with the assumed mechanism behind our training and resonates with the fact that we did not foster word problem solving in the training, but only restructuring additive situation models. We acknowledge that these results need to be backed up with more advanced analyses, e.g., mediation analyses, but we find this result notable.

6.6. *Limitations and outlook*

Overall, the findings support suggestions by Greeno (1980) and Stern (1993) that fostering flexibility in dealing with additive situations by developing the Inversion Strategy and the Dynamization Strategy is a feasible way to develop learners' word problem solving skills. Since the experimental group received additional support in contrast to the control group, it is not possible to draw conclusions about the importance of the training program for regular mathematics lessons, beyond the fact that the training is effective in principle. In the future, the approach should be compared to other approaches to foster word problem solving.

Flexibility in dealing with additive situations was measured with a new test instrument. Although the instrument was sufficiently reliable, it was quite easy and thus did not differentiate well between high-achieving students' performance. The instrument only required receptive flexibility by evaluating the match of two different descriptions, but not to actively produce descriptions statements for the same situation. It is unclear yet, how this productive flexibility can be measured, in particular with this age group and with larger samples. Future research needs to investigate, if and how such instruments can be used to measure students' flexibility and also scrutinize the role of receptive and productive flexibility. Finally, we considered general, and not subject-related language skills as control variable in our study. Our main reason was that it would be hard to disentangle our (already verbal) flexibility construct from subject-related language skills theoretically and empirically, as conceptualizations of subject-related language skills are strongly intertwined with mathematical understanding of the respective concepts on a theoretical level (Ufer & Bochnik, 2020). Future research might address this gap conceptually and empirically, and conceptualize and investigate the relation between flexibility, as a measure strongly focused on understanding how mathematical structures occur in verbal descriptions of real-world situations, and subject-related as well as general language skills.

6.7. *Summary*

Eliciting the application of the Dynamization Strategy and the Inversion Strategy (Greeno, 1980; Stern, 1993) and reflecting on situation structures showed a positive effect on flexibility and word problem solving and thus adds to effective trainings in the field. This observation also underpins the importance of the quality of the students' individual situation model for word problem solving (Leiss et al., 2010; Stern & Lehrndorfer, 1992; Thevenot et al., 2007). It is not plausible that the effect is only caused by the additional instruction, since no actual word problems were solved during the training. Indeed, the study indicates that it is possible to foster word problem solving skills by solely focusing on the level of the situation structure, while working with mathematical operations is (temporarily) left aside. However, further research is needed to explain the mechanisms behind the delayed effects on word problem solving skills and flexibility.

Finally, students benefitted from the training regardless their language skills. This supports the assumption that language-related trainings, including the flexible use of language to describe additive situations, are helpful for all learners, not only for those with low language skills. Moreover, the training provides a language-sensitive approach to support learners with different levels of language skills in word problem solving.

CRediT authorship contribution statement

Laura Gabler: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft. **Stefan Ufer:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2024.101902.

References

Barbu, O. C., & Beal, C. R. (2010). Effects of linguistic complexity and math difficulty on word problem solving by English learners. *International Journal of Education, 2*(2), 1–19. https://doi.org/10.5296/ije.v2i2.508

Blum, W., & Leiß, D. (2007). Deal with modelling problems. In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling: Education, engineering and economics – ICTMA 12* (pp. 222–231). Horwood Publishing.

Bochnik, K. (2017). *Sprachbezogene merkmale als erklärung für disparitäten mathematischer leistung: Differenzierte analysen im rahmen einer längsschnittstudie in der dritten jahrgangsstufe*. Waxmann.

Briars, D. J., & Larkin, J. H. (1984). An integrated model of skill in solving elementary word problems. *Cognition and Instruction, 1*(3), 245–296. https://doi.org/10.1207/s1532690xci0103_1

Cook, S. C., Collins, L. W., Morin, L. L., & Riccomini, P. J. (2020). Schema-based instruction for mathematical word problem solving: An evidence-based review for students with learning disabilities. *Learning Disability Quarterly, 43*(2), 75–87.

Czocher, J. A. (2018). How does validating activity contribute to the modeling process? *Educational Studies in Mathematics, 99*(2), 137–159. https://doi.org/10.1007/s10649-018-9833-4

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*, 1–13. https://doi.org/10.3389/fpsyg.2015.00348

Dumont, H., & Ready, D. D. (2023). On the promise of personalized learning for educational equity. *Npj Science of Learning, 8*(1), 26.

Erath, K., Ingram, J., Moschkovich, J., & Prediger, S. (2021). Designing and enacting instruction that enhances language for mathematics learning: A review of the state of development and research. *ZDM Mathematics Education, 53*(2), 245–262. https://doi.org/10.1007/s11858-020-01213-2

Fromme, M., Wartha, S., & Benz, C. (2011). Grundvorstellungen zur Subtraktion: Tragfähiges Operationsverständnis durch flexible Übersetzungen. *Grundschulmagazin*, (4), 35–40.

Fuson, K. C., Carroll, W. M., & Landis, J. (1996). Levels in conceptualizing and solving addition and subtraction compare word problems. *Cognition and Instruction, 14*(3), 345–371. https://doi.org/10.1207/s1532690xci1403_3

Gabler, L., & Ufer, S. (2020). Flexibilität im umgang mit mathematischen situationsstrukturen: Eine vorstudie zu einem förderkonzept zum lösen von Textaufgaben zu addition und subtraktion [flexibility in dealing with mathematical situation structures: A preliminary study on a training program on solving additive word problems]. *Journal für Mathematik-Didaktik, 42*(1), 61–96. https://doi.org/10.1007/s13138-020-00170-3

Gabler, L., & Ufer, S. (2021). Gaining flexibility in dealing with arithmetic situations: A qualitative analysis of second graders' development during an intervention. *ZDM Mathematics Education, 53*(2), 375–392. https://doi.org/10.1007/s11858-021-01257-y

Gabler, L., & Ufer, S. (2022). Contribution of flexibility in dealing with mathematical situations to word-problem solving beyond established predictors. In *Proceedings of the 45th conference of the international group for the psychology of mathematics education*. Alicante.

Gabler, L., von Damnitz, F., & Ufer, S. (2023). Additive word problems in German 1st and 2nd grade textbooks. In *Proceedings of the 46th conference of the international group for the psychology of mathematics education*. Haifa.

Goulet-Lyle, M.-P., Voyer, D., & Verschaffel, L. (2020). How does imposing a step-by-step solution method impact students' approach to mathematical word problem solving? *ZDM Mathematics Education, 52*(1), 139–149. https://doi.org/10.1007/s11858-019-01098-w

Greeno, J. G. (1980). Some examples of cognitive task analysis with instructional implications. In E. Snow, P.-A. Frederico, & W. E. Montague (Eds.), *2. Aptitude, learning, and instruction* (pp. 1–21). Lawrence Erlbaum Associates. Cognitive process analysis of learning and problem solving.

Hartmann, J., & Fritz, A. (2021). Language and mathematics: How children learn arithmetic through specifying their lexical concepts of natural numbers. In A. Fritz, E. Gürsoy, & M. Herzog (Eds.), *Diversity dimensions in mathematics and language learning* (pp. 21–39). De Gruyter.

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/10.1037/0022-0663.84.1.76

Huang, R., Zhang, Q., Chang, Y.-p., & Kimmins, D. (2019). Developing students' ability to solve word problems through learning trajectory-based and variation task-informed instruction. *ZDM Mathematics Education, 51*(1), 169–181. https://doi.org/10.1007/s11858-018-0983-8

Hudson, T. S. (1980). *Young children's difficulty with "how many more-than... Are there?" Questions*. Dissertation at Indiana University.

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

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review, 24*(2), 313–337. https://doi.org/10.1007/s10648-012-9195-x

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

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

Lein, A. E., Jitendra, A. K., & Harwell, M. R. (2020). Effectiveness of mathematical word problem solving interventions for students with learning disabilities and/or mathematics difficulties: A meta-analysis. *Journal of Educational Psychology, 112*(7), 1388–1408. https://doi.org/10.1037/edu0000453

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

Lenhard, W., & Schneider, W. (2018). *Elfe II: Ein Leseverständnistest für Erst- bis Siebtklässler. Version II*. Hogrefe.

Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79*(4), 363–371. https://doi.org/10.1037/0022-0663.79.4.363

McMullen, J. A., Hannula-Sormunen, M. M., & Lehtinen, E. (2013). Young children's recognition of quantitative relations in mathematically unspecified settings. *The Journal of Mathematical Behavior, 32*(3), 450–460. https://doi.org/10.1016/j.jmathb.2013.06.001

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science, 159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Moschkovich, J. (2015). Academic literacy in mathematics for English Learners. *The Journal of Mathematical Behavior, 40*, 43–62. https://doi.org/10.1016/j.jmathb.2015.01.005

Muth, K. D. (1984). Solving arithmetic word problems: Role of reading and computational skills. *Journal of Educational Psychology, 76*(2), 205–210. https://doi.org/10.1037/0022-0663.76.2.205

Nesher, P., Greeno, J. G., & Riley, M. S. (1982). The development of semantic categories for addition and subtraction. *Educational Studies in Mathematics, 13*(4), 373–394. https://doi.org/10.1007/BF00366618

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

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

Powell, S. R., Driver, M. K., Roberts, G., & Fall, A.-M. (2017). An analysis of the mathematics vocabulary knowledge of third- and fifth-grade students: Connections to general vocabulary and mathematics computation. *Learning and Individual Differences, 57*, 22–32. https://doi.org/10.1016/j.lindif.2017.05.011

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 Mathematics Education, 47*(6), 947–962. https://doi.org/10.1007/s11858-015-0702-7

Prediger, S., & Wessel, L. (2018). Brauchen mehrsprachige Jugendliche eine andere fach- und sprachintegrierte Förderung als einsprachige? *Zeitschrift für Erziehungswissenschaft, 21*(2), 361c382. https://doi.org/10.1007/s11618-017-0785-8

Pyle, D., Pyle, N., Lignugaris/Kraft, B., Duran, L., & Akers, J. (2017). Academic effects of peer-mediated interventions with English language learners: A research synthesis. *Review of Educational Research, 87*(1), 103–133. https://doi.org/10.3102/0034654316653663

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.

Riley, M. S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction, 5*(1), 49–101. https://doi.org/10.1207/s1532690xci0501_2

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

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346–362. https://doi.org/10.1037/0022-0663.93.2.346

Scheibling-Sève, C., Pasquinelli, E., & Sander, E. (2020). Assessing conceptual knowledge through solving arithmetic word problems. *Educational Studies in Mathematics, 103*, 293–311. https://doi.org/10.1007/s10649-020-09938-3

Schipper, W. (2009). *Handbuch für den Mathematikunterricht an Grundschulen*. Schroedel.

Schoenfeld, A. H. (2007). *Assessing mathematical proficiency, 53*. Cambridge University Press.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge in ill-structured domains. *Educational Technology, 31*(5), 24–33.

Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *Journal of Educational Psychology, 85*(1), 7–23. https://doi.org/10.1037/0022-0663.85.1.7

Stern, E. (1994). Die Erweiterung des mathematischen Verständnisses mit Hilfe von Textaufgaben. *Grundschule, 26*(3), 23–25.

Stern, E. (1998). *Die Entwicklung des mathematischen Verständnisses im Kindesalter*. Pabst Science Publishers.

Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word problems. *Cognitive Development, 7*(2), 259–268. https://doi.org/10.1016/0885-2014(92)90014-I

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

Ufer, S., & Bochnik, K. (2020). The role of general and subject-specific language skills when learning mathematics in elementary school. *Journal für Mathematik-Didaktik, 41*(1), 81–117.

Van Lieshout, E. C., & Xenidou-Dervou, I. (2020). Simple pictorial mathematics problems for children: Locating sources of cognitive load and how to reduce it. *ZDM Mathematics Education, 52*(1), 73–85. https://doi.org/10.1007/s11858-019-01091-3

Verschaffel, L. (1994). Using retelling data to study elementary school children's representations and solutions of compare problems. *Journal for Research in Mathematics Education, 25*(2), 141–165. https://doi.org/10.5951/jresematheduc.25.2.0141

Verschaffel, L., Depaepe, F., & Van Dooren, W. (2015). Individual differences in word problem solving. In R. C. Kadosh, & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 953–974). Oxford University Press.

Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics* (pp. 557–628). Information Age Publishers Inc.

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

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

Warner, L. B., Alcock, L. J., Coppolo, J., Jr., & Davis, G. E. (2003). How does flexible mathematical thinking contribute to the growth of understanding? In N. A. Pateman, B. J. Dougherty, & J. Zillox (Eds.), *Proceedings of the 27th conference of the international group for the psychology of mathematics education held jointly with the 25th conference of the national group for the psychology of mathematics education* (pp. 371–378). PME.

Weiß, R. H., & Osterland, J. (2013). *Grundintelligenztest skala 1: Revision; CFT 1-R*. Hogrefe.

Wolters, M. A. (1983). The part-whole schema and artithmetical problems. *Educational Studies in Mathematics, 14*(2), 127–138.