



Gaining flexibility in dealing with arithmetic situations: a qualitative analysis of second graders' development during an intervention

Laura Gabler¹ · Stefan Ufer¹

Accepted: 19 March 2021 / Published online: 4 April 2021
© The Author(s) 2021

Abstract

The influence of language and situation structure on the difficulty of word problems has been investigated intensively in the field of mathematics education. However, instructional approaches to overcoming students' difficulties are still not widely available. This paper describes an intervention to develop second graders' skills in handling additive word problems flexibly. During ten small-group sessions of 40 min, two strategies to restructure the situation described in a word problem were introduced: (1) changing the direction of mathematical relations and (2) changing semantic structures. The introduction of these strategies was supported using macro-scaffolding. The development of students' flexibility in dealing with arithmetic situations during the intervention was analyzed in a longitudinal case study focusing on four students, who were preselected from a larger sample based on their language skills. We examined audio data and student work by applying qualitative content analysis. Students' development in handling word problems flexibly was compared with the intended learning trajectory in the intervention. The results provide insights into potential key processes when gaining flexibility, and yield information on the necessary adaptations of the learning trajectory.

Keywords Additive word problems · Comparison of sets · Macro-scaffolding · Qualitative content analysis

1 Introduction

Solving word problems is difficult for many students. Language plays an important role as learners decode textually represented descriptions of arithmetic situations (Dröse, 2019). These descriptions differ in their linguistic features, such as syntax or semantics, which can influence the difficulty of word problems (e.g., Stern, 1993). To address such difficulties, research has suggested strategies that guide learners to view situations from various perspectives. These strategies are intended to help students construct and reorganize their understanding of the presented situation. Therefore, we describe an intervention program that was intended to enhance students' language use when solving word problems by supporting students in flexibly changing between different views and descriptions of arithmetic situations. To describe situations flexibly, access to words and phrases specific to mathematical concepts ("linguistic

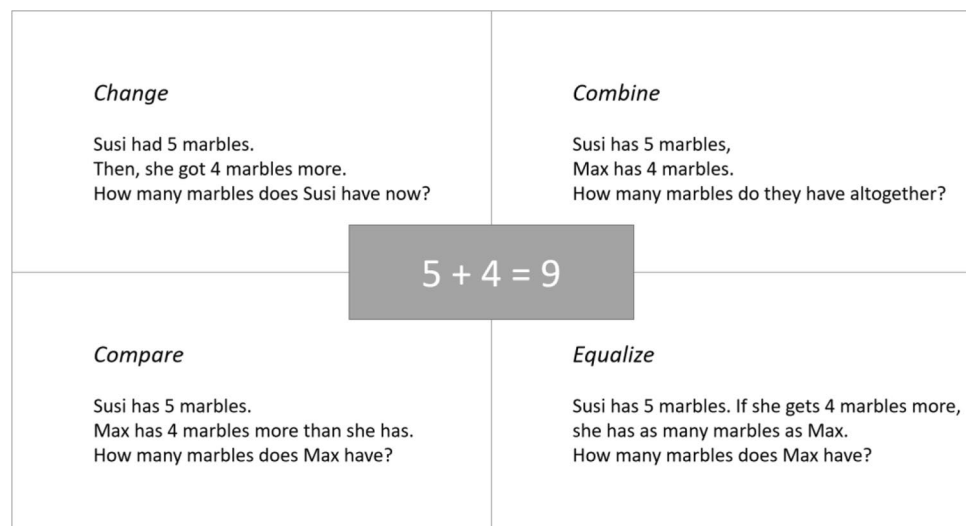
means", Pöhler & Prediger, 2015) seems conducive. Thus, a macro-scaffolding approach offers opportunities to encounter and use such linguistic means.

Following Solano-Flores (2010), we view language not only as a structure (meaning the linguistic features a text presents), but also as a process that focuses on the dynamic aspect of language in mathematical communication and the construction and use of mathematical knowledge. In the intervention, the use of language played a role in the following different ways. (1) Students were encouraged to formulate and transfer between different descriptions of the same arithmetic situation. We intended to support students in using language cognitively to construct accurate and rich mental representations of the given situation (Götze, 2019). (2) To support students with this goal, the language used to describe situation structures was analyzed collaboratively. This was another way to stimulate the cognitive use of language in the program. (3) We integrated the communicative function of language by encouraging rich discourse practices, such as explaining, justifying, or arguing about different descriptions of arithmetic situations (Erath et al., 2021), with the overall goal of supporting students in learning to use language as a tool to construct and organize

✉ Laura Gabler
gabler@math.lmu.de

¹ Department of Mathematics, LMU Munich, Theresienstr. 39, 80333 Munich, Germany

Fig. 1 Semantic structures describing the same mathematical structure



mathematical knowledge in the context of word problems (Sfard, 2008).

It was an open question, if and how learners benefit from these assumed learning opportunities. In this paper, we focus on the development of four selected students during the program. In a qualitative analysis of data from the intervention sessions, we analyzed whether differences in their learning paths could point to parts of the program that are not yet sufficiently adapted to the learners. Starting from classical research on word problems, we outline in the following sections how the intervention is based on suggestions by Greeno (1980), Fuson et al. (1996), and Stern (1993), how the program is designed, and the role language plays in this intervention, followed by a description of a qualitative study involving four students.

2 Theoretical background

2.1 Research on the difficulty of word problems

Previously, the difficulty of additive one-step word problems was analyzed in extensive research (Daroczy et al., 2015; Riley & Greeno, 1988; Stern, 1993). Different features of the presented situation, such as *semantic structure*, *additive or subtractive wording*, and *unknown set*, were found to influence the difficulty of word problems.

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/four types of additive word problems (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*). While *dynamic* word problems (e.g., change) describe actions, combine and compare problems refer to *static* situations. *Equalize* problems, a less common type, combine features of change and compare problems. Past research (e.g., Riley & Greeno, 1988; Stern, 1992) highlighted compare problems as especially difficult types. In compare problems, numbers not only describe concrete sets, but also the difference between two concrete sets, which may be harder to represent mentally (Stern, 1993). Moreover, identifying compared entities and understanding the syntactic structure of the sentence simultaneously (Schleppegrell, 2007) is linguistically demanding. According to Fuson et al. (1996), it is vital to derive from a relational statement *which* quantity is more/less and *how big* the difference between the two quantities is. Indeed, studies by Stern (1993) and Mekhmandarov et al. (1996) indicate that rephrasing compare problems is difficult for many students.

2.1.2 Additive or subtractive wording

Variations in a word problem's wording can also lead to the same mathematical structure. Fuson et al. (1996) distinguished 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, Fig. 1) or 'less', 'smaller' (subtractive wording). 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', Fig. 1) or removing a quantity (subtractive wording, e.g., 'to give away', 'to sell').

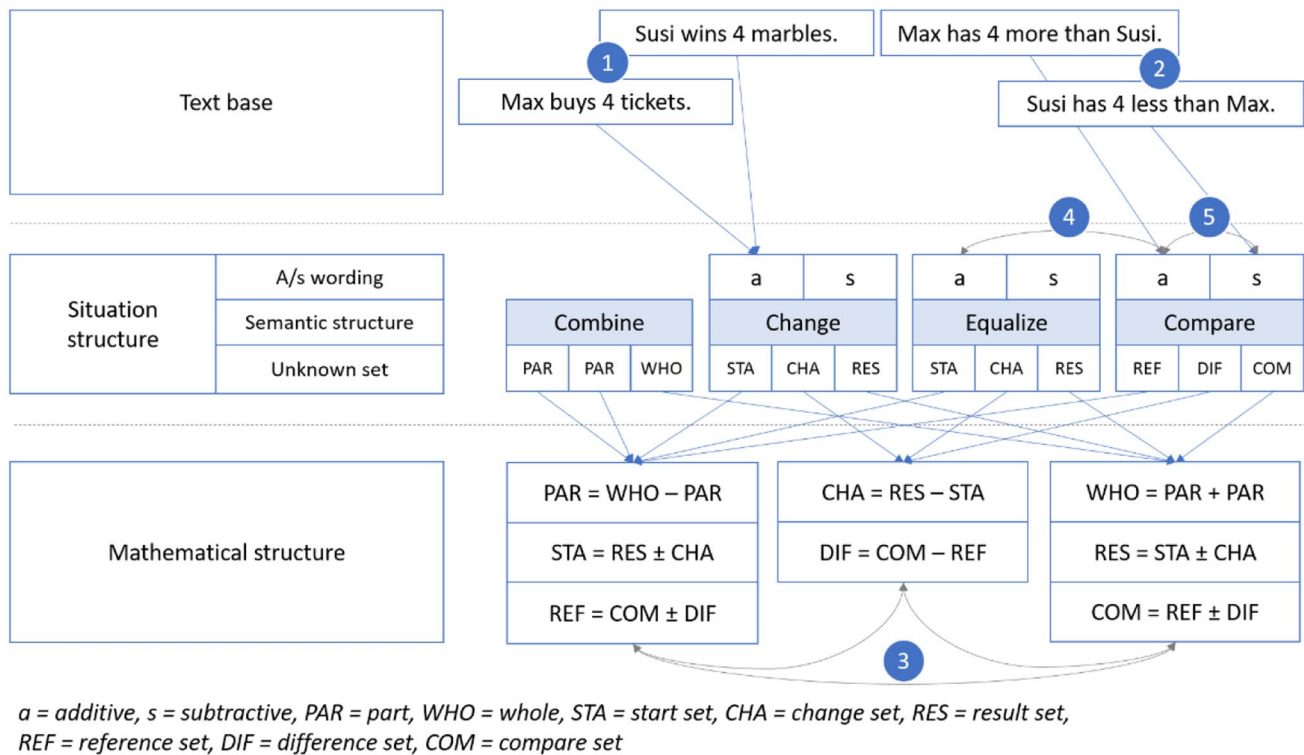


Fig. 2 Model on structure levels of word problems

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*, *difference*, and *compare sets* (e.g., Stern, 1993). Their equivalents in dynamic situations are *start*, *change*, and *result sets*. Studies have shown that word problems with an unknown reference/start set are harder than those with an unknown compare/result set (Riley & Greeno, 1988; Stern, 1992). However, the influence of the unknown set on a word problem’s difficulty is modulated by the a/s wording (Briars & Larkin, 1984; Gabler & Ufer, 2020): Word problems in which the directly applicable mathematical operation¹ (determined by the unknown set, see Fig. 2) is inconsistent with the wording (e.g., subtractive wording but directly applicable addition) are usually harder than consistent word problems (Lewis & Mayer, 1987). Solving inconsistent word problems requires a deep understanding of the situation, since a superficial interpretation (e.g., subtractive wording indicating subtraction) does not lead to a successful solution (Scheibling-Sève et al., 2020).

¹ Even if the directly applicable structure is subtractive, students may transform it into an indirect addition (Fig. 2, ③).

2.2 Flexibility in dealing with arithmetic situations

To overcome the reported barriers in students’ solution processes, Scheibling-Sève et al. (2020) argued that conceptual knowledge about interrelations between problems with different unknown sets, semantic structures, and a/s wording may be helpful. Conceptual knowledge comprises “principles that govern a domain and the interrelations between units of knowledge in this domain” and is expected to help students organize “information in their internal representation of [the] problems” (Rittle-Johnson et al., 2001, pp. 347).

Theories on solving word problems assume that students use this knowledge to construct an individual *situation model* and a *mathematical (problem) model* from the given *text base* (Kintsch & Greeno, 1985). Difficulties can arise from both processes, namely, decoding and transforming the text base into a situation model (“comprehension obstacles”) or describing the individual situation model with corresponding mathematical concepts (“conceptual obstacles”, Prediger & Krägeloh, 2015). The word problem author realizes a specific *situation structure* in the form of a problem text (*text base*) (Fig. 2, Gabler & Ufer, 2020). Different text bases (Fig. 2, example ①) can express the same situation structure, and different formulations of the same situation (Fig. 2, example ②) can highlight different features of the

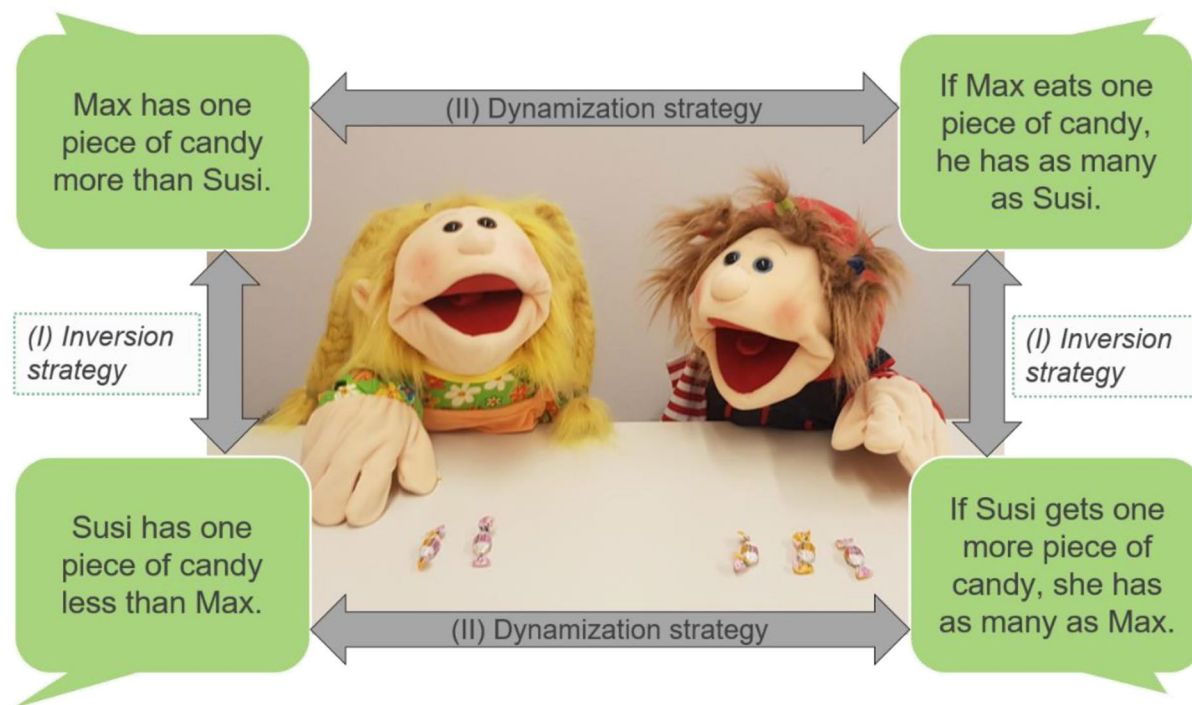


Fig. 3 Examples for inversion and dynamization strategies

situation structure. This intended situation structure corresponds to an intended *mathematical structure*.

With regard to students' perspective, common frameworks (e.g., Kintsch & Greeno, 1985) assume that learners construct individual mental models during the solution process. Learners decode the text base (textual realization of a word problem) into an initial situation model (Kintsch & Greeno, 1985) based on their prior knowledge. At best, this model contains the main components of the intended situation structure. According to models of reading comprehension (Kintsch, 2018), learners can enrich their initial situation model with inferences by adding further features of the situation. For instance, learners could reinterpret compare problems by including an action into their situation model, which equalizes the described difference (Fig. 2, ④, description in Sect. 2.3), or change from an additive structure to a subtractive structure (⑤). In the best case, learners finally transform their individual situation model into a mathematical model, which corresponds to the intended *mathematical structure* of the word problem, or an equivalent one.

Choosing an adequate mathematical operation is contingent on students' conceptual knowledge about connections between situation structures and mathematical structures, but also on which features of a situation structure are included in the students' situation models (Fig. 2). Depending on the reconstructed features of the originally intended situation, it may be more or less straightforward to construct a mathematical model. In this sense, we consider *flexibility*

in dealing with arithmetic situations as a skill to enrich situation models with further structural features. This includes reinterpreting a described situation regarding its situation structure, inferring features of the situation structure that are not described in the text base, and deciding whether a description fits the situation or not.

2.3 Gaining flexibility

Research has suggested the introduction of strategies to reinterpret and enrich situation models with further information (Fuson et al., 1996; Greeno, 1980; Stern, 1993) to enable students to mathematize situation models more easily. We understand strategies as cognitive procedures that have a heuristic value when solving a certain type of problem. We propose two strategies that may lead to the pursued flexibility (Fig. 3), namely, *Inversion* and *Dynamization*.

2.3.1 Inversion strategy: changing the direction of mathematical relations

Stern (1993) and Fuson et al. (1996) stressed the role of a/s wording. 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. Understanding this linguistic *symmetry* may help students solve compare problems (Stern, 1993). Flexible switching between linguistically symmetric

statements (*inverting* the direction of the relational term) may allow students to reinterpret more difficult compare problems with an unknown reference set as empirically easier problems with an unknown compare set (Fig. 2, ⑤, Fig. 3).

2.3.2 Dynamization strategy: changing the semantic structure

Another suggestion is to reinterpret difficult semantic structures as easier ones. Greeno (1980) proposed reinterpreting 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 difficulty of compare problems, we propose to transfer this idea to a similar strategy (Fig. 2, ④, Fig. 3): Students could dynamize compare problems by reinterpreting them as equalize problems, since *dynamic* equalizing may be easier to represent than a static comparison.

Both strategies rely on conceptual knowledge, which is necessary in order to solve word problems (Morales et al., 1985). It helps learners focus on relevant features of the situation structure and add this information to their situation model (Rittle-Johnson et al., 2001). Gaining flexibility may be one way of achieving this result: If learners struggle in solving a difficult word problem, the inclusion of other perspectives on the situation may help them create a more accurate and elaborate situation model. Comparing different descriptions of the same, but structurally different situations, is expected to stimulate learners to make connections between different structural features and enrich their conceptual knowledge of the underlying arithmetic operations.

2.4 Using language to gain flexibility

Understanding and expressing descriptions of situations flexibly is closely related to language skills. Snow and Uccelli (2009) stated that dealing with complex mathematical concepts requires demanding and specific language. *Inversion* builds on well-connected vocabulary on relational terms. *Dynamization* makes use of action verbs and conditional sentences (‘If..., then...’) to express equalization. Following Pöhler and Prediger (2015), introducing the presented strategies must also be accompanied by introducing corresponding content-related linguistic means. One approach to support students with lower language skills is macro-scaffolding. It describes pre-organized support by the teacher, taking into account students’ different language skills (Hammond & Gibbons, 2005), and entails a sequencing of tasks that allows students to progress from accessible to more complex tasks. This scaffolding guides the overall sequencing of learning tasks, but also supports teachers in selecting support *during*

the interaction (so-called micro-scaffolding, Hammond & Gibbons, 2005), for example, by means of visualization or specific language support (Prediger & Pöhler, 2015).

Comparing and contrasting descriptions of situations through Inversion and Dynamization can sensitize students to the linguistic means used in different descriptions and the subtleties of their interpretation. Analyzing these contrasts relates to variation processes, which are connected to the design principle to compare language pieces for raising students’ language awareness (design principle P6, Erath et al., 2021), and to the variation theory promoted by Marton (Pang et al., 2017).

However, language is necessary not only to gain flexibility, but it also works as a tool (Götze, 2019): To develop conceptual knowledge during this analysis of different descriptions, it is advisable to enhance rich discourse practices (design principle P1, Erath et al., 2021). Explaining, justifying, and arguing about descriptions of word problems are typical examples (Moschkovich, 2015), but so also is providing flexible descriptions of a given situation.

In summary, students can benefit from scaffolding in lexical (dealing with linguistic means specific to mathematical concepts) and semantic areas (constructing the meaning of these mathematical concepts) (Prediger & Pöhler, 2015). Since linguistic and content-related demands interact (Kempert et al., 2018), the program combines conceptual and language learning tightly.

3 Intervention design

3.1 Intended learning trajectory and learning tasks

The program is based on an intended learning trajectory (LT). A LT is comprised of a learning goal, specified learning activities, and a hypothetical learning process during these activities (Simon, 1995). The students’ individual learning paths may differ from this intended LT. The intervention consisted of ten 40–50 min small-group sessions led by trained tutors over five weeks. Although the design principles from Erath et al. (2021) were not available in this form during the design of the program, the final design is well aligned with some of them.

Figure 4 shows how the five phases of the program were sequenced over the ten sessions.² After an initial phase of familiarization with certain *Basics* (e.g., comparison statements), Dynamization and Inversion were introduced implicitly over the three main phases of the LT; students first approached the two strategies by *verifying* and *matching*

² More examples of tasks can be found in the supplementary material A.1.

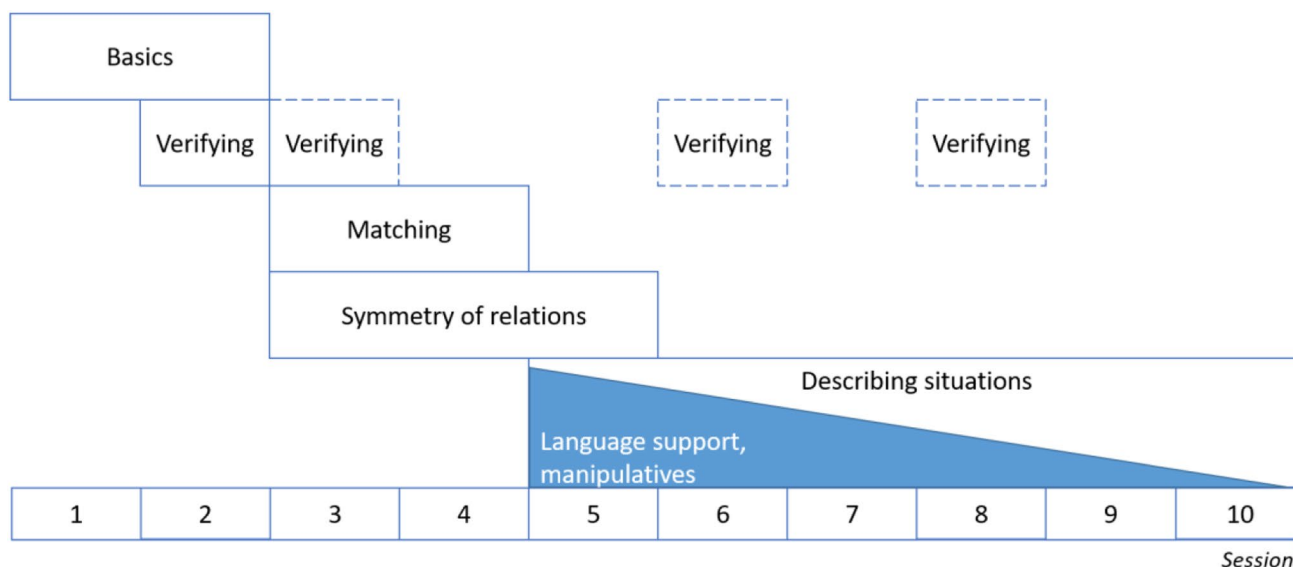


Fig. 4 Procedure of phases during the ten intervention sessions

situations and were then encouraged to transfer to actively *describing* situations in different ways (examples of tasks are presented in Fig. 5). Since researchers emphasize the difficulty of understanding symmetrical relational statements (Sect. 2.3), Inversion was also introduced explicitly (*Symmetry of relations*). Within each phase, tasks were sequenced along the difficulty of situation structures, progressing from situations with unknown compare and difference sets to unknown reference sets (Sect. 2.1). Interrelations between these structures were jointly analyzed to develop conceptual knowledge. By connecting different representations of situations (photos, texts, drawings, manipulatives), we addressed design principle P3 (Erath et al., 2021).

New task types were discussed and solved collaboratively. Tutors shifted responsibility to students continually, with phases of individual and team work followed by joint discussions for most tasks. Language support was faded out as the program progressed. However, tutors provided contingent support within the intervention framework to keep students actively engaged in the tasks.

3.1.1 Basics

To consolidate prior knowledge, we focused on (1) difference sets and (2) equalizing actions. (1) Prior research (e.g., Stern, 1998) highlighted the importance of understanding that two numerical quantities (e.g., 4 and 7) differ by a third quantity (3), and that two quantities cannot only be compared *qualitatively* ('Susi has more'), but that this relation can also be *quantified* ('Susi has 3 more'). When identifying a quantitative comparison, thinking of qualitative comparisons ('Who has more/less?') can support learners' inclusion

of the relation's direction in their situation model (Stern, 1998). If learners do not represent the relation between two sets quantitatively, they may mistake the difference set for a concrete set and interpret a statement such as 'Susi has 3 cards more than Max' as 'Susi has 3 cards' (Mekhmandarov et al., 1996). Thus, tutors were instructed to link questions on qualitative and quantitative comparisons by aiming at qualitative comparisons first and then discussing their quantification.


(2) The Dynamization strategy makes intensive use of equalize situations (Sect. 2.3). Since these situations rarely occur in textbooks, they require clarification. Before the students worked with equalize situations, we discussed how equalizing relates to the manipulation of *one* set instead of *both* sets. Students played the game *Hamstern* (Verboom, 2010), which provided a context to discuss compare and equalize statements in the same situation (see A.1).

3.1.2 Verifying and matching

Before describing situations actively, *given* statements on arithmetic situations were discussed and contrasted. This provided learners with linguistic means for the flexible description of compare and equalize situations. Both phases contained variations in the statements linked to the two strategies (Fig. 3). Analyzing and comparing these statements encouraged the use of language for knowledge organization by emphasizing interrelations between different descriptions of situations.

(1) In Sessions 2 and 3, students *verified* given statements on situations about two different quantities (Fig. 5). The

Verifying *Which statements are true? Explain how you know.*



✓

Susi has two cards less than Max.

Susi has two cards more than Max.

Max has two cards more than Susi.


Max has two cards less than Susi.

✗

Matching *Which statement fits which picture? Explain how you know.*

Susi has two cards less than Max.

Susi has two cards more than Max.




Max has two cards more than Susi.

Max has two cards less than Susi.

Describing *What do you know about the situation?*

They are playing cards.

Susi has two cards and Max has four.

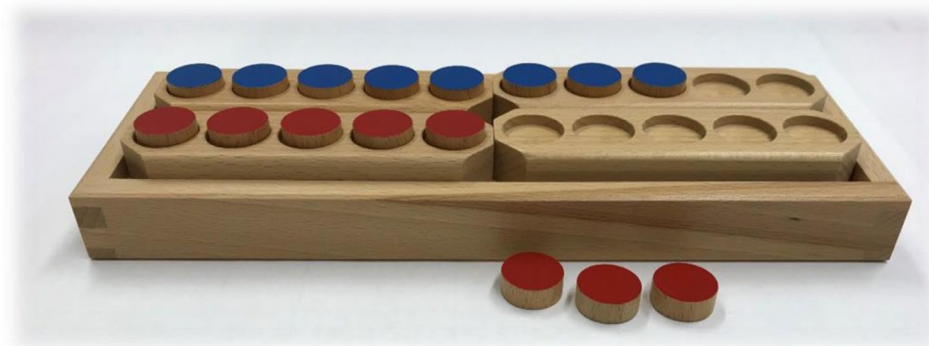


Max has two cards more than Susi.

If Susi takes two cards, she has as many as Max.

Fig. 5 Examples of tasks for the three main phases of the intervention

Fig. 6 Rechenschiffchen



students decided whether a statement matched a picture of a situation and justified their decisions (design principle P1, “Enhance rich discourse practices”, Erath et al., 2021). To adapt to the intervention’s progress in difficulty, situations with unknown compare and reference sets were included in Sessions 6 and 8. Over Sessions 3, 6, and 8, Verifying tasks were also used to track students’ individual progress systematically.

- (2) In Sessions 3 and 4, students *matched* statements to two situations with swapped concrete sets; for instance, Susi had two cards more than Max in one picture and vice versa (Fig. 5). By contrasting statements on these inverse situations, this phase was intended to systematize students’ experiences with descriptions of compare and equalize situations they had gathered during Verifying. Moreover, contrasting statements on these inverse situations highlighted the linguistic subtleties, in order to raise students’ language awareness (design principle P6, Erath et al., 2021). During all Matching tasks, tutors were instructed to enhance arguing, why certain statements match a situation and how they differ (design principle P1, Erath et al., 2021), to establish structured but adaptable mathematics language routines (design principle P2), and intensify students’ experiences with linguistic means for compare and equalize situations.

3.1.3 Symmetry of relations

During the explicit introduction of Inversion, students learned to invert relational statements. After freely formulating relational statements triggered by providing relational terms, tutors supported the students to first invert qualitative and then quantitative comparison statements. Tutors chose support means from the intervention script from light to strong support (see A.1). Furthermore, linguistic means were addressed by dealing with expressions helpful for describing the *a/s* wording.

3.1.4 Describing situations

Learners were asked to actively articulate descriptions of given situations (Fig. 5). By formulating varying descriptions, we encouraged learners to use language cognitively to enrich their situation models. The students analyzed descriptions of situations in multiple modes (listening, talking, reading, writing) and communication settings. Rich discourse practices were enhanced by encouraging flexible descriptions of arithmetic situations and explanations of the differences and commonalities between different descriptions and the presented situations (design principle P1, Erath et al., 2021).

Students received two types of support, as follows. (1) Language support: Tutors provided incomplete sentence templates that gave a rough structure for students’ own descriptions. This scaffold was removed gradually, until the students could describe situations with a focus on comparison and equalizing without support. For flexible language production, tutors were also instructed to supply word cards to trigger comparison (e.g., ‘more’, ‘less’) or equalizing (e.g., ‘If..., then...’). (2) Manipulatives: Students also visualized situations with *Rechenschiffchen*, a common teaching manipulative in German classrooms similar to twenty frames (Fig. 6).

The direct, visual comparison of sets was assumed to highlight the one-to-one correspondence and part-whole relationships, and to activate conceptual knowledge on additive word problems (Morales et al., 1985). Students were also asked to equalize sets using the *Rechenschiffchen* to build up mental representations of equalizing actions and compared sets. Tutors encouraged verbalizing thoughts and actions when working with the *Rechenschiffchen*. Later, the support by the *Rechenschiffchen* was faded out in order to establish independence from manipulatives.

3.2 The role of language in the intervention design

Traditional research on additive word problems has strongly focused on identifying language-related features that influence their difficulty (e.g., Daroczy et al., 2015). The consequence cannot be to avoid these features in instruction. It is vital to find ways to support students in understanding harder problem types. The intervention design took into account the role of language in achieving this in different ways, as follows.

3.2.1 Cognitive function of language during word-problem solving

The program encouraged students to rephrase harder problem types into easier ones. This strategy, proposed by Greeno (1980) and Stern (1993), targets the cognitive use of language *during word-problem solving*. Our goal was not to simplify harder word problem types by rewording them beforehand (e.g., Vicente et al., 2008), but to amplify students' language use to re-interpret them in simpler situation structures (Erath et al., 2021; Schleppegrell, 2007).

3.2.2 Cognitive function of language during learning

Students find this strategy helpful only if they are sensitive to the ways in which different descriptions of a situation are related to each other. We assumed that students can enrich their conceptual knowledge regarding such situations by connecting these different descriptions into a network of linked perspectives on arithmetic situations. In this vein, we expected that analyzing how language is used in different ways to describe situations would provide fruitful learning opportunities. We consider this an example of using language cognitively when *learning* about the structural features of such situations (Götze, 2019).

3.2.3 Communicative function of language during learning

Finally, to achieve this analysis, we encouraged students to think of varying descriptions of the same situation, to reason why these descriptions fit the same situation, and to discuss structural similarities and differences. The aim of this aspect was to use *communication* for learning processes during the intervention (Moschkovich, 2015).

4 Research questions

Based on the research on word problems and the role of language in mathematics learning, we describe an intervention to foster students' flexibility in dealing with arithmetic situations. In this paper, we aim to analyze the development

of four selected students' flexibility and we focus on the following questions:

4.1 Q1: Which differences in students' learning paths point to parts at which the intended LT is not sufficiently adapted to individual students?

We assumed that students would make use of the provided learning activities in different ways. Investigating such differences, we aimed to discover typical patterns and systematic obstacles when students gained the pursued flexibility. These patterns and obstacles may highlight potential 'key processes' that require special attention when supporting students during the LT.

4.2 Q2: How does students' flexibility develop during the intervention?

Considering the potential key processes from Q1, we investigated the students' ability to deal flexibly with arithmetic situations, and how this ability changed during the intervention. We expected progress regarding the strategies, but were also interested in finding out if specific aspects, such as dealing with compare situations, would be harder to develop for some students.

5 Design and method

5.1 Context and case sampling

The program was conducted in ten classes in three schools in Germany. Within each class, we selected three students with higher and lower language skills, each based on the ELFE II reading test (Lenhard & Schneider, 2018), to form ten intervention groups of six students each ($N=60$ in total). The ELFE II test provides a rough assessment of students' language proficiency based on reading speed and accuracy. The intervention was conducted in a separate room, predominantly during German language lessons. Pre-service teachers acted as tutors and were instructed beforehand. The intervention script, content and procedure, duration of the phases, wording suggestions, and use of student support were discussed. For comparability, tutors followed a specified sequence with determined options to adapt to the individual needs of students. In a pilot intervention with $N=4$ students from another school, we examined the suitability of the tasks beforehand.

For the qualitative analysis, we selected four of the sixty students before the start of the intervention based on pre-test data. Since support means were primarily targeted at students with limited language proficiency, we selected pairs

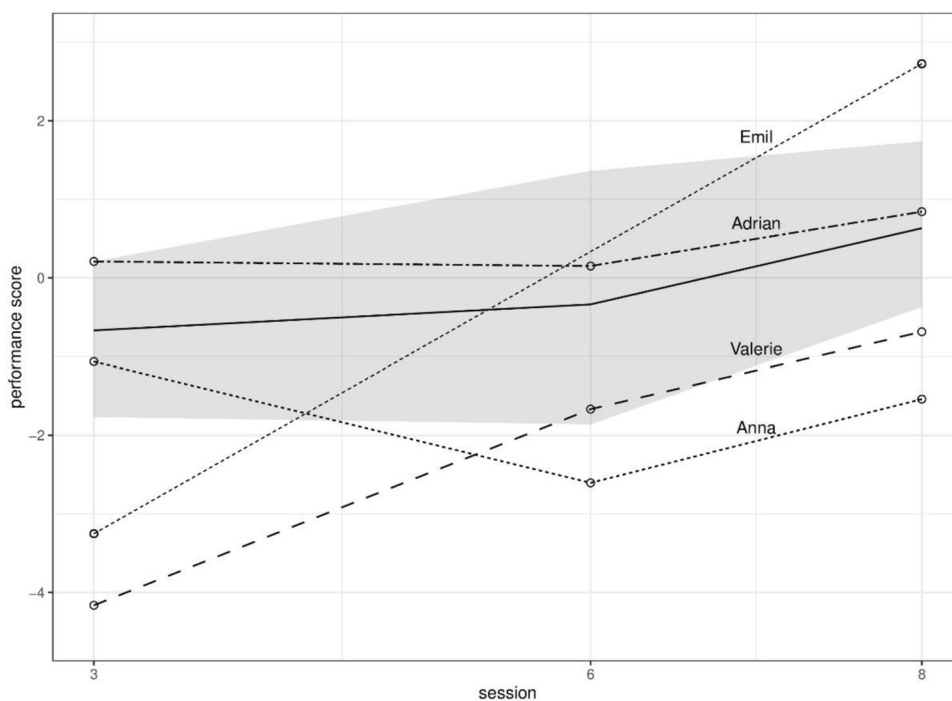
Table 1 Student profiles

Subject	Gender	ELFE II ^a (0–100)	Basic arithmetic skills ^b (0–1)	Languages spoken at home
Valerie	w	9.70	0.50	Croatian, German
Anna	w	21.20	0.38	Serbian, German
Adrian	m	8.10	0.25	German, Italian
Emil	m	21.20	0.50	German

^aPercentile rank with respect to the norm sample of the ELFE II reading test

^bPercentage of correct answers

Fig. 7 Students' development in Verifying over Sessions 3, 6, and 8 (solid line indicates intervention sample mean; shaded area between 20 and 80% quantile)



of students with lower reading test scores. The four selected students came from two different intervention groups, which were instructed by the same tutor (Group 1: Valerie,³ Anna. Group 2: Adrian, Emil; Table 1).

Their basic arithmetic skills were at the lower (Anna, Adrian) and lower average levels (Valerie, Emil). While Valerie and Anna predominantly spoke a language different from the instruction language at home, Adrian and Emil spoke mostly or exclusively German at home. Due to illness, Emil missed Sessions 5 and 6.

5.2 Overview of students' development

We first provide a rough overview of the development of the entire intervention sample ($N=60$) and the four selected students. Sessions 3, 6, and 8 included Verifying worksheets,

which were linked pairwise by common tasks. Students worked on these worksheets individually, and their ideas were discussed jointly afterward. Initial responses were scored dichotomously, and linked performance scores for students' flexibility were calculated for each student and session using the 1PL Rasch model (Rasch, 1960).⁴ Figure 7 displays the sample mean and the sample mean plus/minus one standard deviation (see also supplementary material A.2) of students' performance scores by session (solid lines). A repeated measures ANOVA indicated that the intervention was successful because a significant average progress of the intervention sample over the three sessions ($F(121.92, 2) = 18.94, p < 0.001, \eta_p^2 = 0.21$) could be observed.

Figure 7 also shows the performance scores of the four selected students. The development of students' performance

³ The students' names were changed for data protection.

⁴ The scale was anchored by setting the latent mean of person scores over all sessions to 0.

Table 2 Coding manual

Category	Code	Exemplary statement
(1) Verbalizing comparison		
Qualitative comparison	COM-QUAL	'Susi has more marbles'
Quantitative comparison		
Concrete set	CON	'Susi has 7 marbles'
Difference set	DIF	'Susi has 4 marbles more than Max'
(2) Verbalizing equalization	EQ	'Max needs to get 4 marbles to have as many as Susi'
(3) Changing the semantic structure		
From comparison to equalization	COM-DYN	'Susi has 4 marbles more than Max' → 'When Susi puts 4 marbles away, she has as many marbles as Max'
From equalization to comparison	EQ-DYN	'When Susi puts 4 marbles away, she has as many marbles as Max' → 'Susi has 4 marbles more than Max'
(4) Changing the direction of mathematical relations		
Compare situations	COM-INV	'Susi has 4 marbles more than Max' → 'Max has 4 marbles less than Susi'
Equalize situations	EQ-INV	'When Susi puts 4 marbles away, she has as many marbles as Max' → 'When Max takes 4 more marbles, he has as many marbles as Susi'

scores differed substantially, some (e.g., Adrian) developing roughly parallel to the sample mean, others showing substantial progress (Emil) or even a slight decrease in performance (Anna). The standard errors of individual performance estimates were between 0.57 and 1.60, showing that a reliable quantitative analysis of individual students' development in Verifying was not possible and qualitative analysis was needed to gain deeper insights.

5.3 Method: qualitative content analysis

For the qualitative analysis, all intervention sessions were recorded and transcribed. These transcripts and the students' worksheets were investigated following the principles of qualitative content analysis (Mayring, 2014). This category-based approach is characterized by a strong orientation in guiding research questions (Mayring, 2014). Each student's statement was counted as one coding unit. Phases of group work were omitted because contributions could not be attributed to single students. We started with a theory-based coding manual to identify different manifestations of flexibility and adapted it during the analysis. Table 2 shows the final coding manual. The first two categories addressed formulating comparison and equalizing statements, which are considered important prerequisites for gaining flexibility (Sect. 3.1). Due to the specific difficulty of compare situations, additional subcategories were included. The third and fourth categories reflected the two strategies (Sect. 2.3). Whenever a student formulated a comparison or equalizing statement, and *then* immediately applied the Dynamization or Inversion strategy, the answer was coded as category 3 or 4.

Each statement in a single task was coded separately. For each statement, we coded also which answer the respective task required. An extra code was used exclusively for open questions with more than one possible answer type. Coding was conducted by two independent raters. The results indicate a very good interrater reliability ($\kappa = 0.85$) (Landis & Koch, 1977).

We analyzed the development over sessions for each student and finally contrasted these paths between different students. This process was followed repeatedly with different emphases (Fig. 8) arising from the background and the design of our intervention. We started from the coded data and proceeded to the raw data to check and enrich our initial interpretations. In addition, we followed perspectives arising from noteworthy observations in the raw data, in order to reconsider each student's development.

6 Results

6.1 Q1: Uncovering potential key processes

Three major differences in students' learning paths emerged during the analyses (key processes, KP), which may point to parts of the LT that were not yet sufficiently adapted to individual students.

6.1.1 (KP1) Distinguishing concrete and difference sets

One major difference in the students' learning paths emerged in the ways they interpreted difference sets in comparison

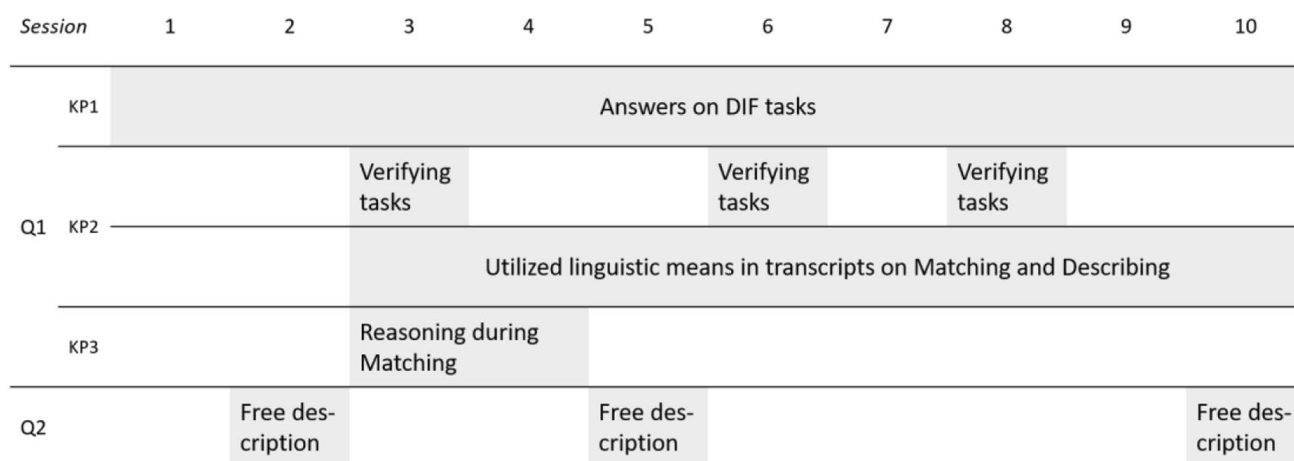


Fig. 8 Overview of different emphases during the data analysis, subdivided by research questions (Q1, Q2) and emerging key processes (KP1, KP2, KP3)

Table 3 Sets mentioned by students during tasks that required determining a difference set (DIF = difference set, CON = concrete set)

Session	Valerie		Anna		Adrian		Emil	
	DIF	CON	DIF	CON	DIF	CON	DIF	CON
1	0	8	7	0	6	0	9	0
2	3	0	3	0	4	0	1	0
3	4	1	4	1	7	0	3	2
4	0	0	0	0	2	1	0	0
5	1	1	1	0	5	0	Absent	
6	6	1	7	2	8	0	Absent	
7	2	2	5	2	5	0	6	0
8	3	4	3	4	7	0	6	1
9	0	0	0	2	0	0	0	0
10	0	1	1	0	3	0	1	0
Average per session	1.9	1.8	3.1	1.1	4.7	0.1	3.25	0.38

statements. To investigate these differences, we considered student answers on tasks aimed at determining difference sets (Fig. 8). Mostly, these answers were either related to difference [DIF] or concrete [CON] sets. In Table 3, we compare how often each student named a difference or concrete set when answering such tasks for each session.

Adrian and Emil mostly referred to difference sets and seemed to interpret comparison statements correctly from the beginning (Table 3). Valerie and Anna, however, mentioned concrete sets instead of difference sets frequently and throughout the program. This result indicates that they did not fully benefit from the corresponding learning opportunities in the Basics phase. During the program, they made some, but slower, progress in this aspect.

An explanation for their slower progress could be that they often seemed to understand statements such as ‘There are 7 sheep more than cows’ as two messages: ‘There are 7 sheep’ [CON] and ‘There are more sheep than cows’ [COM-QUAL]. They seemed to link numbers to concrete sets, and relational statements separately to qualitative comparison.

For example, during Verifying in Session 3, two concrete sets were given: ‘There are 7 sheep and 4 cows’. When Anna worked on the statement ‘There are 7 sheep more than cows’, Anna classified this statement as correct. Her answer indicates that she interpreted the qualitative relationship correctly (more sheep than cows). However, she seemed to identify the numerical information (7) as a concrete set instead of a difference set. This observation is supported by the following excerpt:

Table 4 Correct answers for each given situation (represented as picture or text) containing six subitems each (DIF = unknown difference set, COM = unknown compare set, REF = unknown reference set)

Given situation	Session 3		Session 6			Session 8		
	DIF	DIF	DIF	DIF	COM	DIF	COM	REF
Representation	Picture	Text	Picture	Text	Text	Text	Text	Text
Valerie	4/6	5/6	6/6	4/6	3/6	5/6	6/6	1/6
Anna	6/6	5/6	5/6	4/6	1/6	4/6	4/6	1/6
Adrian	6/6	5/6	6/6	6/6	5/6	6/6	5/6	5/6
Emil	6/6	1/6	<i>Absent</i>			6/6	6/6	6/6

Session 5, group 6

The given situation is read aloud: ‘There are 3 apples and 5 bananas in the bowl’. The tutor asks students to integrate the words ‘more’ and ‘less’.

Anna: Well, there are less apples [COM-QUAL] and more bananas [COM-INV] in the bowl.
 T⁵: Okay, can you express that more precisely?
 Anna: In the bowl, there are... 3 apples...
 Valerie: I know it, too!
 Anna: ... 5 bananas... [CON]

After Anna successfully described the situation with relational terms [COM-QUAL] and even inverted this qualitative relation [COM-INV], the tutor encouraged her to quantify the relation. Instead, she named the given concrete sets [CON]. This might further indicate that Anna linked numbers rather to concrete sets, and relational statements separately to qualitative comparison. Similarly, Valerie referred to concrete sets in comparison statements systematically when answering a worksheet on the game *Hamstern* in Session 1, although the tutor had supported her previously in interpreting the difference set by contrasting the sets verbally. Indeed, she determined the difference set correctly at this point:

Session 1, group 6

The students play Hamstern with the tutor. After determining who has more chips, the tutor encourages Valerie to quantify the difference.

T: Valerie, what do you think, how many do I have more [DIF]?
 Valerie: Valerie: You have 6 [CON].
 T: I do have 6, but how many do I have more than you? Think about it.
 Valerie: 4 [DIF].

⁵ T = Tutor.

T: 4, exactly. So, how many [chips] am I allowed to take [EQ]?
 Valerie: 4 [EQ].

In Session 2, she distinguished concrete and difference sets correctly during a similar worksheet. However, Valerie still seemed to struggle with this distinction occasionally until the end of the program.

Although difficulties in understanding difference sets were anticipated and thus considered in the LT, we did not expect them to appear as systematically and frequently as it was the case with Valerie and Anna. Thus, the difficulties could be tackled only partially in the Basics phase and were not fully resolved until the end. It is plausible that such difficulties in the beginning would limit students’ chances to profit from further parts of the program.

6.1.2 (KP2) Transferring from verifying to matching and describing situations

Successful handling of Verifying tasks was assumed to be a helpful activity to stimulate flexible descriptions. In Sessions 3, 6, and 8, we observed whether students had already gained initial flexibility. Table 4 shows how they progressed differently on the three situation types. Adrian gave only a few unsystematic wrong answers throughout the three sessions. Since Emil misread the given situation with unknown difference set in Session 3, he answered almost all respective tasks incorrectly (Table 4). Due to his reading mistake, he assigned the concrete sets to the wrong persons in the situation. All his answers were correct, given this alternative situation model. In Session 8, Emil answered all items correctly.

Valerie and Anna developed Verifying skills later and did not reach the same level as Adrian and Emil. Situations involving two concrete sets (DIF) were easier for them than other types (Table 4). When verifying statements on situations with unknown compare set (COM, Session 6), both showed insecurities initially. Especially Anna seemed to struggle in interpreting the qualitative relation of sets (who has more?). In Session 8, both students showed substantial growth, indicating that they included difference sets in their situation model. However, they still struggled with unknown

reference sets (REF), which might indicate that the program should provide better opportunities for them to develop conceptual knowledge.

As intended in the LT, the students' ability to verify statements advanced from situations occurring in simpler word problem types to more difficult ones over the whole intervention. However, Valerie and Anna progressed more slowly on tasks presenting relational sets verbally (situations with unknown compare/reference sets). Progress in Verifying should prepare students for Matching and Describing tasks by providing linguistic means and encouraging learners to use language for knowledge organization (Sect. 3).

Since language support should encourage this transfer from comprehension to active production of descriptions, we decided to analyze the use of language support in this transition more closely. Transcripts on Matching and Describing tasks were investigated with a specific focus on instances where language support (word cards, sentence frames, sentence starters) was used (Fig. 8). Emil and Adrian had few problems describing arithmetic situations quite early in the program. For Valerie and Anna, the tutor offered more language support in formulating suitable statements. Valerie had problems in formulating an equalizing statement in Session 5. She succeeded in determining the change set, but struggled to complete the sentence with an action verb:

Session 5, group 6

- Elisa: *[reads aloud the provided sentence frame]* If I ..., then my tower is as tall as yours.
- T: What should she do? Valerie. Do you remember, what we did there?
- Valerie: If I one... eh? From Sebastian?
- T: So, try to think about it again.
- Valerie: If I one, then... at this tower... as tall as yours.
- T: If you do what? 'Then my tower is as tall as yours is'.
- Valerie: If I...
- T: What can you do, so that the tower is as tall as this one?
- Valerie: One away?
- T: Exactly! Let's do that.
- Valerie: If I one... away... if I... eh?
- Elisa: I know! If I one, then...
- T: You need more words.
- Elisa: If I one brick... then...
- T: What do you do with the brick?
- Elisa: If I take one brick away, then my tower is as tall as yours.
- T: Fine, do that, Elisa, and now let's check if it's true... is the tower as tall as hers now?
- Students: Yes!

T: Okay, let's put the brick back.

[The next sentence frame is provided]

Valerie: I know!

T: Okay, Valerie, you can try, you already started so well before.

Valerie: If I add one, then it is as tall as yours.

Another student (Elisa) took over to help, and elaborated a possible description with the tutor's help. Following this example, Valerie managed to describe equalization in the next task. Subsequently, her vocabulary on action verbs expanded continually. This indicates that the transfer from Verifying to more advanced parts of the intended LT cannot be taken for granted. Individual language support seemed to be of particular importance for Valerie and Anna when gaining flexibility. In this case, a combination of the sentence frame and the support by her peer allowed Valerie to progress in describing equalizations.

6.1.3 (KP3) Reasoning with comparisons when matching statements

The Matching phase revealed differences in the ways students explained why certain statements or pictures were similar or different. To investigate these differences, we contrasted answers with the codes [CON] and [DIF] for such tasks (Fig. 8). While Adrian and Emil reasoned with comparison statements frequently (Adrian, 4 × DIF; Emil, 5 × DIF), Valerie and Anna mostly used the given concrete sets (Valerie, 8 × CON; Anna, 2 × CON, 2 × EQ). The two excerpts following the question 'What is the difference between the two pictures?' illustrate these observations:

Session 3, group 6

- Valerie: In this picture, Max has 2 [CON], and Susi has 4 [CON].
- T: Exactly.
- Valerie: And here, Max has 4 pieces [CON], and Susi has 2 cards [CON].
- T: Exactly.
- Valerie: Maybe, because they exchanged their cards?

It seems that, at this point, Valerie does not use relations to contrast the situations. This complies with the tendency to mostly link numbers to concrete sets (KP1). In a previous task on the same situation, she indeed matched a description to the wrong picture based on this tendency. Anna matched an equalizing statement on the same situation to the correct picture later, but also referred only to concrete sets in her explanation. Adrian and Emil used comparison statements to contrast the situations and applied the Inversion strategy.

Table 5 Tasks involving the free description of situations in Sessions 2 and 5 (orally in a joint conversation) and Session 10 (in written form as individual work)

Session	Valerie	Anna	Adrian	Emil
2	1 CON	1 CON	1 CON 1 DIF	1 CON
5	1 CON	1 EQ 1 EQ-INV	1 DIF 1 COM-INV 1 EQ 1 EQ-INV	<i>Absent</i>
10	3 CON 1 DIF 3 EQ 2 EQ-INV	3 CON 5 COM-QUAL 1 DIF 5 COM-INV	2 COM-QUAL 3 DIF 5 COM-INV 2 COM-DYN 1 EQ-INV	1 DIF 1 COM-INV 1 COM-DYN 3 EQ 4 EQ-INV

Session 3, group 5

Adrian: Here, Max has 2 more [DIF], and there, Susi has 2 more [DIF].
 T: Exactly. Can you say that in other words, Emil?
 Emil: Here, Susi has 2 less [COM-INV], and there, Max has 2 less [COM-INV].

It seems that enhancing such discourse practices can unveil students' conceptual knowledge and their perception of situation structures. Here, providing useful terms for alternative descriptions (e.g., 'more') could have triggered comparison statements and strengthened Valerie and Anna's awareness of relations. Tutors were instructed to use such relational terms as triggers in predetermined situations, but were not prepared to draw on this kind of support spontaneously.

6.2 Q2: Overall development of flexibility

Considering these key processes, we examined the four students' overall development of flexibility. Three tasks for actively describing situations were selected to provide insights into students' development (Fig. 8). During these tasks, the students were encouraged to describe situations without explicit instruction orally in a group setting (Sessions 2 and 5) and individually in written form (Session 10). This procedure should reveal whether they formulated varying descriptions and used language cognitively to enrich their situation models.

For the four students, all with lower language skills, we found different developmental patterns. Adrian was the only one to formulate comparison statements spontaneously already in Session 2. Despite very low language skills and comparably low arithmetic pre-test scores, he quickly adopted the two strategies. In Session 5, he added

descriptions on equalizing and Inversion systematically and Dynamization in Session 10 (Table 5). In line with a consistent focus on relations, he preferred formulating comparison statements. The following excerpt illustrates Adrian's systematic approach of describing situations flexibly:

Session 7, group 5

T: Let's have a look at this picture...
 Adrian: Can I start?
 T: ...and describe what we see.
 Adrian: Can I start?
 T: Adrian. Their names were Susi and Max.
 Adrian: Max has 4 keys [CON], Susi has 2 keys [CON].
 T: Please show that to me at the *Rechenschiffchen*. Susi is red [*color of chips*]. [...] So, and now...
 Adrian: Max has 2 more than Susi [DIF]. Susi has 2 less than Max [COM-INV]. If Susi gets 2 keys more, then they're as many keys... they're as many keys...
 T: As Max's.
 Adrian: ...as Max's [COM-DYN]. If Max puts 2 keys away, then they're as many as Susi [EQ-INV].

Starting with higher pre-test language and mathematics scores than Adrian, Emil first focused on concrete sets in Session 2 (Table 5). Although he missed two sessions, Emil adopted both strategies and gained flexibility with a strong focus on equalizing statements until Session 10. He and Adrian required little language support beyond what was offered by sequencing from Verifying over Matching to Describing tasks and the corresponding language support (KP2). Moreover, both distinguished concrete and difference sets since early sessions (KP1).

However, Valerie connected numbers almost exclusively to concrete sets during the whole program (KP1). When reasoning about situations, she mostly focused on concrete sets as well (KP3). It seems that overcoming this issue would have required a stronger focus on difference sets and comparison statements or more adaptive language support. This is most likely a reason for her slower progress in Verifying tasks beyond those with unknown difference set. Given the LT's structure, problems in Verifying probably made it hard for her to work on further tasks meaningfully, and the tutor struggled to support her effectively. As a result, her progress regarding flexibility was small: Session 10 reveals signs of progress when she named not only concrete sets, but also one (incorrect) statement on the difference set and several equalizing statements and their inversions (Table 5). For Valerie, adaptive deviation from the LT might have been promising. Indeed, students' reasoning (KP3) seems to provide indications if such adaptations are warranted.

Anna started out with higher pre-test language performance, but with a slightly weaker mathematics performance

than Valerie. Similarly, she linked numbers primarily to concrete sets at first (KP1). However, she made progress in the key process. Like Valerie, she struggled with Verifying tasks beyond those involving two concrete sets (KP2), and her explanations indicated that a stronger focus on difference sets would have been helpful (KP3). While she focused on equalizing in Session 5, she tried to formulate comparison statements and their inversion in Session 10 (Table 5). However, she then applied the Inversion strategy incorrectly: After writing a correct statement such as ‘There are more nuts and less mandarins [COM-QUAL, COM-INV]’, she also wrote down the opposite of the given situation: ‘There are more mandarins and less nuts [COM-QUAL, COM-INV]’. It seemed as if Anna did apply Inversion, but did not focus on describing the *same* situation. Furthermore, slower initial progress might have kept her from benefiting from subsequent activities that should prepare learners to gain such strategies.

7 Discussion

Initial analysis indicated an overall progress in flexibility for the entire intervention sample. Qualitative analysis with four selected learners matched this trend and extended these findings, yielding three potential key processes for the successful development of flexibility in dealing with arithmetic situations. From these observations, we first draw conclusions on parts of the intended LT that require specific attention to address a wider range of learning paths (Q1):

KP1: Interpreting relations as a quantitative phenomenon and using numbers to describe difference sets seems to be a key process in gaining flexibility. This points to a very specific interpretation of comparison statements, but it shows how important the development of a sound understanding of linguistic means is, in order to describe situations from various perspectives. It also illustrates how conceptual knowledge of different situation structures is closely connected to understanding the linguistic means describing these structures (Pöhler & Prediger, 2015). Following Barwell (2005), it is vital to find learning tasks that contrast opposing interpretations in a learning group, in order to build a common understanding. This aspect could be addressed using Verifying tasks that contrast situations differentiating between statements such as ‘3 sheep more than cows’ and ‘3 sheep, and more sheep than cows’ in future revisions of the intervention (Mekhmandarov et al., 1996).

KP2: As expected, transferring linguistic means from Verifying to Matching and Describing tasks was feasible. We assume that not only encountering linguistic means in the Verifying phase, but also discussing the use of these linguistic means, are crucial (design principle P1, Erath et al., 2021; Moschkovich, 2015). This assumption is supported by

the observation that language support by the tutor was vital for the transition to Describing. This observation underpins how important explicit support (Hammond & Gibbons, 2005) can be to help students make use of linguistic means when reflecting on situation structures (Sect. 3.2).

KP3: Our observations reveal the power of language to uncover students’ conceptual knowledge and flexibility regarding situation structures (communicative function of language). Learners focusing on specific semantic structures or concrete sets in their descriptions might indicate that other situation structures should be discussed more intensively with the learner. By asking students to explain differences between situations, teachers could utilize this discourse practice to investigate which learning opportunities can encourage students to enrich their situation models and gain access more easily to both strategies.

Second, the analysis provided information on the students’ development of flexibility (Q2) and, consequently, students’ cognitive use of language during word-problem solving. Some students progressed mostly as intended. Others made progress along the intended LT, but took substantially more time. This result supports the assumption that the approaches proposed by Stern (1993) and Greeno (1980) are useful for fostering some students’ flexibility (Gabler & Ufer, 2020). However, individual learning paths differed in several key processes, to which the learning opportunities were not yet sufficiently adaptive. While all four students progressed on equalize problems substantially, their progress varied more in the compare situations. However, the fact that Anna and Valerie made some progress shows that initial problems do not necessarily imply that flexibility cannot be gained during the intervention. The differences rather seem to be derived from a primarily qualitative interpretation of comparison statements. However, other factors, such as prior knowledge, may also be the basis of different learning paths. Adrian gained substantial flexibility, despite having lower language skills. It seems that he already had a tendency to focus on quantitative relations (McMullen et al., 2013), which might have given him a good starting point for using the two strategies. Future research may investigate reasons for variation in students’ learning by considering individual and didactical aspects more systematically.

Although qualitative analysis can uncover relevant aspects that would be hidden from a more summative, quantitative approach, our study must be viewed in the light of some limitations. It cannot be ruled out that motivational aspects or their mathematical self-concepts influenced students’ development. Our analyses were restricted to four pre-selected students. To substantiate our results, we sampled transcripts repeatedly and contrasted the cases, searching for additional data against which we could test our interpretations. Moreover, the newly created LT requires further modification (Simon, 1995) and more adaptive learning

activities to meet the learners' needs better. However, the intervention's determined structure helped contrast learners' paths reliably.

Despite these limitations, the analyses show that fostering the pursued flexibility is generally possible, but goes along with substantial and possibly systematic heterogeneity. The results concerning students' development endorse the feasibility of the chosen approach to support students in constructing richer, more accurate situation models and provide a starting point to address students' difficulties with word problems by enhancing language. Encouraging the cognitive and communicative use of language has turned out to be an expedient approach to achieving this goal. It still needs to be investigated if amplifying language (Schleppegrell, 2007) with this instructional approach is helpful for learners during actual word-problem solving. Moreover, only quantitative analyses can clarify whether a substantial part of the participating group of students could gain flexibility in the program and if they transferred this skill to compare problems.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11858-021-01257-y>.

Funding Open Access funding enabled and organized by Projekt DEAL.

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

References

- Barwell, R. (2005). Ambiguity in the mathematics classroom. *Language and Education, 19*(2), 117–125.
- Briars, D. J., & Larkin, J. H. (1984). An integrated model of skill in solving elementary word problems. *Cognition and Instruction, 1*(3), 245–296.
- 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.
- Dröse, J. (2019). *Textaufgaben lesen und verstehen lernen: Entwicklungsforschungsstudie zur mathematikspezifischen Leseverständnisförderung*. (Vol. 43). Springer.
- 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).
- 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.
- Gabler, L., & Ufer, S. (2020). Flexibilität im Umgang mit mathematischen Situationsstrukturen: Eine Vorstudie zu einem Förderkonzept zum Lösen von Textaufgaben zu Addition und Subtraktion. *Journal für Mathematik-Didaktik, 42*(1), 61–96.
- Götze, D. (2019). The cognitive function of language and its influence on the learning of mathematics in inclusive settings: A primary school study on the example of multiplication and division. In D. Kollosche, R. Marcone, M. Knigge, M. Penteado, & O. Skovsmose (Eds.), *Inclusive mathematics education*. (pp. 357–375). Springer.
- Greeno, J. G. (1980). Some examples of cognitive task analysis with instructional implications. In E. Snow, P.-A. Frederico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction. Volume 2: Cognitive process analysis of learning and problem solving*. (pp. 1–21). Lawrence Erlbaum Associates.
- Hammond, J., & Gibbons, P. (2005). Putting scaffolding to work: The contribution of scaffolding in articulating ESL education. *Prospect, 20*(1), 6–30.
- Kempert, S., Schalk, L., & Saalbach, H. (2018). Übersichtsartikel: Sprache als Werkzeug des Lernens: Ein Überblick zu den kommunikativen und kognitiven Funktionen der Sprache und deren Bedeutung für den fachlichen Wissenserwerb. *Psychologie in Erziehung und Unterricht, 56*, 1–20.
- Kintsch, W. (2018). Revisiting the construction-integration model of text comprehension and its implications for instruction. In D. E. Alvermann, N. J. Unrau, M. Sailors, & R. B. Ruddell (Eds.), *Theoretical models and processes of literacy*. (pp. 178–203). Routledge.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*(1), 109–129.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.
- Lenhard, W., & Schneider, W. (2018). *ELFE II: Ein Leseverständnistest für Erst- bis Siebtklässler*. 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.
- Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. <https://nbn-resolving.org/urn:nbn:de:0168-ssao-395173>
- 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.
- Mekhmandarov, I., Meron, R., & Peled, I. (1996). Performance and understanding: A closer look at comparison word problems. *PME20*, Valencia.
- Morales, R. V., Shute, V. J., & Pellegrino, J. W. (1985). Developmental differences in understanding and solving simple mathematics word problems. *Cognition and Instruction, 2*(1), 41–57.
- Moschkovich, J. (2015). Academic literacy in mathematics for English Learners. *The Journal of Mathematical Behavior, 40*, 43–62.
- Pang, M. F., Bao, J., & Ki, W. W. (2017). 'Bianshi' and the variation theory of learning: Illustrating two frameworks of variation and invariance in the teaching of mathematics. In R. Huang & Y. Li (Eds.), *Teaching and learning mathematics through variation*. (pp. 43–67). Sense Publishers.
- Pöhler, B., & Prediger, S. (2015). Intertwining lexical and conceptual learning trajectories: A design research study on dual macro-scaffolding towards percentages. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(6), 1697–1722.

- Prediger, S., & Krägeloh, N. (2015). Low achieving eighth graders learn to crack word problems: A design research project for aligning a strategic scaffolding tool to students' mental processes. *ZDM*, *47*(6), 947–962.
- Prediger, S., & Pöhler, B. (2015). The interplay of micro-and macro-scaffolding: An empirical reconstruction for the case of an intervention on percentages. *ZDM*, *47*(7), 1179–1194.
- Rasch, G. (1960). *Studies in mathematical psychology: I.* . 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.
- 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.
- Scheibling-Sève, C., Pasquinelli, E., & Sander, E. (2020). Assessing conceptual knowledge through solving arithmetic word problems. *Educational Studies in Mathematics*, *103*, 293–311.
- Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, *23*(2), 139–159.
- Sfard, A. (2008). Thinking as communicating: Human development, the growth of discourses, and mathematizing. *TMME*, *5*, 429–436.
- Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*(2), 114–145.
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy*. (pp. 112–133). Cambridge University Press.
- Solano-Flores, G. (2010). Function and form in research on language and mathematics education. In J. Moschkovich (Ed.), *Language and mathematics education: Multiple perspectives and directions for research*. (pp. 113–149). Information Age.
- Stern, E. (1992). Warum werden Kapitänsaufgaben „gelöst“? Das Verstehen von Textaufgaben aus psychologischer Sicht. *Der Mathematikunterricht*, *38*(5), 7–29.
- 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.
- Stern, E. (1998). *Die Entwicklung des mathematischen Verständnisses im Kindesalter.* . Pabst Science Publishers.
- Verboom, L. (2010). Ich habe 3 Plättchen mehr als du. *Grundschule-Mathematik*, *25*, 6–7.
- Vicente, S., Orrantia, J., & Verschaffel, L. (2008). Influence of situational and mathematical information on situationally difficult word problems. *Studia Psychologica*, *50*(4), 337–356.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.