Flexibility in dealing with additive word problems

Theoretical foundation, development, and evaluation of an intervention program

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

eingereicht von

Laura Gabler

10. August 2021

1. Gutachter/in: Prof. Dr. Stefan Ufer (Ludwig-Maximilians-Universität München)

2. Gutachter/in: Prof. Dr. Silke Ruwisch (Leuphana Universität Lüneburg)

3. Gutachter/in: Prof. Dr. Birgit Neuhaus (Ludwig-Maximilians-Universität München)

Tag der mündlichen Prüfung: 7.12.2021

Danksagung

Ich möchte mich herzlich bei allen bedanken, die mich bei dieser Dissertation begleitet und unterstützt haben.

An erster Stelle möchte ich meinem Doktorvater Prof. Dr. Stefan Ufer für die großartige Betreuung danken. Danke, dass Sie mir diese Arbeit ermöglicht und stets konstruktiv, verlässlich und optimistisch betreut haben. Darüber hinaus bedanke ich mich für all die Möglichkeiten und Einblicke, die mir während der Promotion gewährt wurden.

Ebenfalls danken möchte ich Prof. Dr. Silke Ruwisch für das hilfreiche Feedback und die Begutachtung der Arbeit. Außerdem danke ich Prof. Dr. Birgit Neuhaus für die Begutachtung dieser Dissertation und Prof. Dr. Markus Heydenreich für die Übernahme des Prüfungsvorsitzes bei der Disputation.

Ganz besonders möchte ich mich auch bei meinen Kolleginnen und Kollegen am Lehrstuhl der Didaktik der Mathematik der LMU bedanken. Die interessanten, lustigen und oft auch tiefgehenden Gespräche mit euch haben meine Promotion zu einer unvergesslichen Zeit gemacht. Ich bedanke mich bei Katrin, Jin, Patricia, Alex W., Anastasia, Helga, Kathrin N., Frau Ekimova, Daniel, Maike, Kathrin K., Alex R., Tanja, Luzia, Dani, Tina, Simon, Sarah, Steffi, Constanze, Matze und Timo.

Im Speziellen möchte ich Katrin danken – danke, dass du mich auf den Weg zur Promotion gebracht und meine Begeisterung zum Thema Mathematik und Sprache geweckt hast. Patrizia danke ich für die herzliche Aufnahme als Büropartnerin zu Beginn. Kathrin N. möchte ich für ihren Scharfsinn, ihre aufmerksamen Fragen und ihre wertvollen Ideen zur Förderung danken. Tanja danke ich für ihre Kreativität, ihren Witz und die schönen, langen Gespräche. Für die unerschöpfliche Hilfsbereitschaft, zum Beispiel bei der Vorbereitung auf die Eignungsprüfung, und für das gut durchdachte, gewissenhafte Feedback, aber auch für die vielen lustigen Momente möchte ich mich bei Daniel ganz herzlich bedanken. Auch Timo gilt mein besonderer Dank für seine Hilfsbereitschaft und die stetige Lust, sich in fremde Themen einzudenken, aber auch für die schöne Zeit gemeinsam im Büro. Insbesondere möchte ich Sarah für ihre herzliche Art, die vielen Lacher und die guten Ratschläge danken. Darüber hinaus freue ich mich, dass über die Zeit am Lehrstuhl hinaus eine Freundschaft fürs Leben entstanden ist.

Mein Dank gilt auch allen Schulen, Lehrkräften, Schülerinnen und Schülern und Hilfskräften, die am Projekt beteiligt waren und damit die Umsetzung der Studien erst ermöglicht haben.

Außerdem danke ich all den Studierenden, die ihre Zulassungsarbeit im Rahmen des Dissertationsprojekts geschrieben haben. Besonders möchte ich Felicitas für ihr besonderes Engagement, ihre Begeisterung und ihre Gedanken zum Thema bedanken.

Zu guter Letzt bedanke ich mich für die Unterstützung im privaten Umfeld. Ich danke meiner Familie für die aufbauenden Worte (und Pakete) und das offene Ohr. Ganz besonderer Dank gilt meinem Freund Rupi für die Zuversicht von Beginn an, für den so wichtigen, wohltuenden Ausgleich und für den unermüdlichen Rückhalt.

Summary

Word problems have been the subject of extensive research for decades (Morales, Shute, & Pellegrino, 1985; Stern & Lehrndorfer, 1992; Verschaffel, Schukajlow, Star, & Van Dooren, 2020). This research paid particular attention to different types of additive word problems and differences in their difficulty. Many studies identified the word problem's situation structure as a major factor influencing its difficulty (Cummins, Kintsch, Reusser, & Weimer, 1988; De Corte & Verschaffel, 1987; Stern, 1998). Tightly connected to this observation is the role of language in word-problem solving (Peng et al., 2020), since word problems represent arithmetic situations verbally (Verschaffel, Greer, & De Corte, 2000). Therefore, language-sensitive instructional approaches to support students dealing with difficult word problems are needed. Indeed, researchers have suggested strategies to help learners overcome such difficulties (Greeno, 1980; Stern, 1993). According to these suggestions, learners could reorganize their situation model by integrating different perspectives on the depicted situation. However, corresponding interventions are still not widely available and rarely investigated.

The dissertation project addresses this research gap by considering the ability to integrate different perspectives into the situation model as a new ability construct (*flexibility in dealing with arithmetic situations*). A special characteristic of this construct is that it completely relies on the level of the situation and thus gets by without conducting mathematical operations. The main goal of this project was to investigate, if fostering students' flexibility in dealing with arithmetic situations can be a feasible approach to support students with difficult additive word problems.

Before an intervention study on fostering the pursued flexibility could be designed, a preliminary study was conducted to clarify two issues. On the one hand, most of the findings on the difficulty of additive word problems date back to the eighties and nineties. Thus, the long-standing research field was revisited by replicating and systematizing these prior findings within a preliminary study in second grade (N = 139). Although the study could replicate prior findings only partly, findings still revealed persisting differences in difficulty of additive word problems and thus the necessity to conceptualize corresponding interventions. On the other hand, the preliminary study investigated, if students already make use of flexibility in dealing with arithmetic situations spontaneously, when they encounter structurally similar word problems. Since it was not observed that students already made use of similar structures in this setting, it was investigated, if an explicit training to develop the pursued flexibility would help learners with solving difficult word problems.

Consequently, an intervention was conceptualized guided by the proposed strategies (Greeno, 1980; Stern, 1993) and a hypothetical learning trajectory (Simon, 1995). Ten grade 2 classrooms (N = 113) participated at an experimental intervention study gathering information

on students' flexibility and word-problem solving skills in pretest, posttest, and follow-up test. The pretest also collected information on students' language skills, general cognitive abilities, basic arithmetic skills and knowledge, and their socio-economic status as control variables. After the pretest, six second graders were selected from each classroom based on their language skills and took part at ten small-group sessions to develop flexibility in dealing with arithmetic situations. The remaining students served as a control group to facilitate the comparison of both groups' development through the intervention. Since the intervention study served as a "feasibility study", the students of the experimental group received the training *additionally* to regular math lessons. The analyses provided first indications that the intervention is a feasible way to foster the pursued flexibility and shows a positive long-term effect on word-problem solving skills. Regarding the influence of language skills, all learners seemed to benefit equally from the intervention. This supports the assumption that language skills, including the flexible use of language to describe arithmetic situations, play an important role in mathematics learning, and that corresponding language support is helpful for all students, not only for those with lower language skills.

To gain more detailed insights into the effects of the intervention, a qualitative content analysis (Mayring, 2014) of the small-group intervention sessions was conducted with preselected students (N = 4) from the intervention group. The observations support the findings that the chosen approach is a feasible way to foster students' flexibility and provide insights into students' development. In particular, they point to necessary adaptations of the hypothetical learning trajectory to address a wider range of learning paths.

In summary, this dissertation project contributes a new ability construct to the field, which was conceptualized based on existing theories. The project supplies a method how flexibility in dealing with arithmetic situations can be measured and instructional approaches how conceptual and language learning can be intertwined to foster students' flexibility. Finally, the analyses provide empirical evidence that fostering students' flexibility with the chosen instructional approach is not only feasible, but also supports learners with solving additive word problems. These findings can serve as a starting point to further investigate the construct of flexibility in dealing with arithmetic situations and how successful activities from the intervention could be integrated into regular math lessons effectively.

Zusammenfassung

Seit Jahrzehnten wird das Lösen von Textaufgaben intensiv beforscht (Morales et al., 1985; Stern & Lehrndorfer, 1992; Verschaffel et al., 2020). Besondere Aufmerksamkeit wurde dabei den Schwierigkeitsunterschieden verschiedener Typen additiver Textaufgaben gegeben. Zahlreiche Studien identifizierten die Situationsstruktur einer Textaufgabe als einen zentralen Einflussfaktor auf die Aufgabenschwierigkeit einer Textaufgabe (Cummins et al., 1988; De Corte & Verschaffel, 1987; Stern, 1998). Eng damit verbunden ist die Rolle von Sprachkompetenzen beim Lösen von Textaufgaben (Peng et al., 2020), da in diesen arithmetische Situationen verbal repräsentiert werden (Verschaffel et al., 2000). Zur Unterstützung von Lernenden beim Lösen schwieriger Textaufgaben sind Instruktionsansätze notwendig, in welchen sprachliches und konzeptuelles Lernen verknüpft wird (Pöhler & Prediger, 2015). In der Tat wurden in der bisherigen Literatur (Greeno, 1980; Stern, 1993) bereits verschiedene Strategien vorgeschlagen, mit deren Hilfe Lernende schwierige Textaufgaben besser lösen könnten. Diesen Vorschlägen zufolge könnten Lernende ihr individuelles Situationsmodell umstrukturieren, indem sie verschiedene Perspektiven auf die dargestellte Situation in ihr Situationsmodell integrieren. Entsprechende Interventionen wurden bisher jedoch kaum entwickelt und beforscht.

Das Dissertationsprojekt knüpft an diese Forschungslücke an und betrachtet die Fähigkeit, verschiedene Perspektiven in ein Situationsmodell zu integrieren, als ein neues Fähigkeitskonstrukt (*Flexibilität im Umgang mit arithmetischen Situationen*). Ein spezielles Merkmal dieses Konstrukts ist der ausschließliche Fokus auf die Situationsebene, ohne dabei auf die Arbeit mit mathematischen Operationen zurückzugreifen. Das Hauptziel des Projekts war zu ermitteln, ob die Förderung von Flexibilität im Umgang mit arithmetischen Situationen einen möglichen Ansatz zur Unterstützung von Lernenden beim Lösen von schwierigen additiven Textaufgaben darstellt.

Vor der Konzeption einer Interventionsstudie zur Förderung der angestrebten Flexibilität wurde eine Vorstudie durchgeführt. Das Ziel war zum einen, ältere Ergebnisse zur Schwierigkeit von additiven Textaufgaben auf ihre Aktualität zu prüfen und zu systematisieren. Dies wurde mit einer Stichprobe von $N = 139$ Kindern der zweiten Jahrgangsstufe untersucht. Obwohl die Vorstudie frühere Befunde nur zum Teil replizieren konnte, zeigten sich nach wie vor bestehende signifikante Schwierigkeitsunterschiede zwischen den verschiedenen Typen additiver Textaufgaben. Dies weist erneut auf die Notwendigkeit von Instruktionsansätzen zur Unterstützung von Lernenden hin. Zum anderen sollte in der Vorstudie untersucht werden, ob Lernende bereits von ihrer potentiell vorhandenen Flexibilität im Umgang mit arithmetischen Situationen spontan Gebrauch machen, wenn ihnen strukturell ähnliche Textaufgaben zur Verfügung gestellt werden. Da solch ein spontaner Gebrauch in diesem Rahmen nicht

beobachtet wurde, stellte sich die Frage, ob ein explizites Training Lernende zur Entwicklung und Nutzung von Flexibilität anregen könnte.

Folglich wurde eine Intervention basierend auf den vorgeschlagenen Strategien (Greeno, 1980; Stern, 1993) und eines antizipierten Lernweges (Simon, 1995) entwickelt. Eine experimentelle Interventionsstudie mit zehn zweiten Klassen (N = 113) erfasste die Flexibilität und die Fähigkeit der Lernenden, Textaufgaben zu lösen, in einem Vortest, einem Nachtest und einem Follow-up-Test, um den Leistungszuwachs in beiden Skalen abzubilden. Zudem wurden Sprachkompetenzen, kognitive Grundfähigkeiten, arithmetische Basisfertigkeiten und der sozioökonomische Status der Lernenden als Kontrollvariablen erhoben. Im Anschluss an den Vortest wurden sechs Lernende pro Klasse anhand ihrer Sprachkompetenzen ausgewählt und innerhalb von zehn Kleingruppensessions gefördert. Die übrigen Lernenden fungierten als Kontrollgruppe, sodass beide Gruppen hinsichtlich deren Leistungszuwachs über die drei Messzeitpunkte hinweg verglichen werden konnten. Da die Intervention als „Machbarkeitsstudie" angelegt war, wurden die Lernenden der Experimentalgruppe *zusätzlich* zum regulären Mathematikunterricht gefördert. Die Analysen weisen darauf hin, dass das entwickelte Förderkonzept in der Tat eine Möglichkeit darstellt, die angestrebte Flexibilität zu fördern und zeigen einen positiven langfristigen Effekt auf die Fähigkeit der Lernenden, Textaufgaben zu lösen. Bezüglich des Einflusses von Sprachkompetenzen schienen alle Lernenden gleichermaßen von der Förderung zu profitieren. Dies stützt die Annahme, dass Sprachkompetenzen, eingeschlossen der flexible Einsatz von Sprache zur Beschreibung arithmetischer Situationen, eine wichtige Rolle bei der Entwicklung mathematischer Kompetenzen spielen. Darüber hinaus weist dies darauf hin, dass sprachliche Unterstützung nicht ausschließlich für Lernende mit niedrigeren Sprachkompetenzen, sondern für *alle* Lernenden, ungeachtet ihrer Sprachkompetenzen, hilfreich sein kann.

Zur genaueren Untersuchung der Wirksamkeit der Intervention wurde eine qualitative Inhaltsanalyse (Mayring, 2014) der Fördersitzungen durchgeführt. Dafür wurden transkribierte Tonaufnahmen mithilfe eines Kodierschemas analysiert, um die Entwicklung von im Vorfeld ausgewählten, sprachlich schwächeren Lernenden ($N = 4$) im Detail abzubilden. Die Beobachtungen unterstützen die Ergebnisse der quantitativen Analysen, dass der gewählte Ansatz zur Förderung von Flexibilität im Umgang mit arithmetischen Situationen geeignet ist, und liefern detailliertere Einblicke in die Entwicklung der Lernenden. Insbesondere weisen die Beobachtungen auf notwendige Anpassungen des Förderkonzepts hin, sodass in Zukunft noch vielfältigere Lernwege angemessen unterstützt werden können.

Zusammenfassend platziert dieses Dissertationsprojekt basierend auf bestehenden Theorien ein neues Fähigkeitskonstrukt im Feld der Mathematikdidaktik. Das Projekt generiert eine Methode zur Messung von Flexibilität im Umgang mit arithmetischen Situationen und darüber hinaus Instruktionsansätze, wie konzeptuelles und sprachliches Lernen zur Förderung von

Flexibilität verknüpft werden kann. Schließlich liefern die Analysen empirische Evidenz, dass die Förderung von Flexibilität mit den gewählten Instruktionsansätzen nicht nur möglich ist, sondern Lernende auch beim Lösen von additiven Textaufgaben unterstützt. Diese Ergebnisse können als Ausgangspunkt für weitere Forschung genutzt werden, um das Konstrukt der Flexibilität im Umgang mit arithmetischen Situationen weiterzuentwickeln und die Wirksamkeit entsprechender Lerngelegenheiten im Regelunterricht zu untersuchen.

Table of contents

1 Introduction

Language skills clearly influence students' development of mathematics skills. According to current discussions, this influence can be partly explained by the mechanism that learners use language for thinking when building mathematics skills (Kempert, Schalk, & Saalbach, 2019; Sfard, 2008). In particular, students use language as a tool to construct and organize mathematical knowledge (Maier & Schweiger, 1999). This tool is assumed to be particularly important in the context of word-problem solving, since learners deal with mathematical structures, which are represented verbally in the problem text (Dröse, 2019; Peng et al., 2020). Many studies reported that word-problem solving is a particularly challenging task for primary school students (for an overview, see Daroczy, Wolska, Meurers, & Nuerk, 2015). Therefore, language-sensitive instructional approaches are needed to support learners with these challenges.

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, Voyer, & Verschaffel, 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. Flexibility is assumed to be essential when dealing with new, unfamiliar situations (Warner, Alcock, Coppolo Jr., & Davis, 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 is what 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." Flexible thinking is also promoted by curricula: In the profile for the subject mathematics, the Bavarian curriculum *LehrplanPLUS* emphasizes the importance of interconnected thinking[1]. In the US-American context, the Standards for Mathematical Practice (National Governors Association Center for Best Practices, 2010) promote flexibility in "using different properties of operations and objects" in problem situations. Researchers already suggested specific strategies, which encourage exactly such flexibility in the context of word-problem solving (Greeno, 1980; Stern, 1993). These strategies build on the idea that learners can add further perspectives to a word problem and interpret the described situation as a familiar, more accessible word problem. Such strategies may be used as the foundation for an intervention program that aims at developing flexibility in dealing with the arithmetic situations described in word problems.

[1] This is worded as "vernetzt denken" in the profile for mathematics in primary schools in the LehrplanPLUS, Section 1.3.

This dissertation project takes up these suggestions and investigates, if fostering flexibility in dealing with arithmetic situations by providing the suggested strategies may enhance students' understanding of difficult word problems. The literature review begins with outlining the role of language when learning mathematics in general (Chapter 2.1). Word-problem solving is then introduced as a content area in mathematics, in which language is particularly prominent. In Chapter 2.2, it is described, which types of word problems are investigated in this work and which processes occur during word-problem solving. Chapters 2.3 and 2.4 provide an overview, which individual features (e.g., language skills of the learner) and task features (e.g., linguistic complexity of the text) play a role in word-problem solving. Finally, theoretical and empirical implications to support students in dealing with challenging task features are discussed and the approach to foster students' flexibility in dealing with arithmetic situations is outlined (Chapter 2.5). Based on the literature review, Chapter 3 specifies the research gap and the associated goals of the dissertation project. The subsequent two chapters report on two studies: a preliminary study (Chapter 4), which prepared further research, and an intervention study (Chapter 5). In this intervention study, it was analyzed how fostering the ability to add further perspectives to word problems and reinterpret them as easier word problems influences the students' flexibility in dealing with such word problems and to what extent this helps with solving difficult word problems. After introducing the intervention study (Chapter 5.1) and its design (Chapter 5.2), a quantitative (Chapter 5.3) and a qualitative analysis (Chapter 5.4) of data from the intervention study are presented. Findings from these two studies are discussed in respect of their contribution to the field, open questions, and implications for future research and teaching (Chapter 6).

In this dissertation, the findings from the project have been compiled for the first time. Parts of the dissertation were already published in two journal articles (Gabler & Ufer, 2020, 2021). In the *Journal für Mathematik-Didaktik* (Gabler & Ufer, 2020), the framework and results from the preliminary study were published in a similar form in German language (parts of Chapters 2 and 4). The co-author of this publication advised the author of this dissertation as far as the design and analysis of this preliminary study are concerned. Further, the co-author provided intensive feedback during the writing and revision process. The article in the journal *ZDM – Mathematics Education* (Gabler & Ufer, 2021) reports on the intervention's design and the qualitative analysis of the intervention study in a similar form in English language (parts of Chapters 2, 5.2, and 5.4). The co-author of this publication consulted the author of this dissertation regarding the study's and the intervention's design. In addition, the co-author provided intensive feedback during the writing and revision process. The main work regarding these publications was accomplished by the author of this dissertation. Regarding both publications, the author organized and implemented the studies, conducted the analyses, and generated the first drafts of the manuscripts.

2 Literature review

2.1 Using language for learning mathematics

When learners engage with mathematics, language plays a role in their learning process: Learners *read* mathematical texts when working on tasks, *listen* to teachers and peers as language models (e.g., explaining a calculation path, logical reasoning), *write* down solution paths and answers during practice and test situations, and *speak* about mathematics when participating in classroom discourse. Accordingly, language skills can be viewed as an overarching construct of being proficient in receptive (reading, listening) and productive (writing, speaking) language use (Jude, 2008).

This chapter outlines how language is related to learning mathematics[2] and thereby lays the groundwork for understanding the role of language when dealing with the verbal representation of arithmetic situations (e.g., in the form of word problems). After reporting on the findings of several studies, which investigated the relation between language and mathematics skills (Chapter 2.1.1), there will be an overview of situations in the classroom, in which language is used for learning mathematics. Potential mechanisms explaining the relation between language and mathematics skills will be presented (Chapter 2.1.2). Finally, the characteristics of language needed for learning mathematics will be specified in greater detail (Chapter 2.1.3).

2.1.1 The relation between language and mathematics skills

Mathematics has been perceived widely as a subject that is based on symbols and not necessarily associated with language. At least since the early 2000s, investigations within national and international large-scale studies, such as the *Programme for International Student Assessment* (PISA) and the *Internationale Grundschul-Lese-Untersuchung* (IGLU, international reading assessment in primary schools), have changed this image fundamentally by proving a positive correlation between reading and mathematics skills (Baumert et al., 2001; Bos et al., 2003). In a meta-analysis of 344 studies, Peng et al. (2020) identified a moderate relation between language and mathematics (r = .42). These findings provide first indications that language and mathematics skills are tightly connected.

Besides language skills, other factors may influence mathematics performance. For a long time, researchers mainly investigated the role of students' backgrounds on mathematics performance. For example, large-scale studies confirmed repeatedly that learners with a migration background are particularly disadvantaged in this respect (e.g., OECD, 2013; Tarelli, Schwippert, & Stubbe, 2012). Other studies focused on the role of family language[3] and found

[2] *Learning mathematics* is used in this dissertation as a term to describe the development of mathematics skills.

[3] *Family language* refers to the language(s) that a person primarily speaks at home with their family.

a similar pattern for learners with a family language different from the language of instruction (e.g., Heinze, Herwartz-Emden, & Reiss, 2007; Prediger, Wilhelm, Büchter, Gürsoy, & Benholz, 2018). In Germany, these differences in mathematics performance of learners with non-German and German family languages already exist in the first school year (Heinze et al., 2007). Cross-sectional analyses reported that language skills can explain differences depending on the family language to a large extent (e.g., Prediger et al., 2018; Ufer, Reiss, & Mehringer, 2013). Longitudinal analyses confirmed this for overarching (Ufer et al., 2013) as well as differentiated measures of language skills (for vocabulary: Paetsch, Felbrich, & Stanat, 2015; for reading comprehension skills: Paetsch, Radmann, Felbrich, Lehmann, & Stanat, 2016) as significant predictors of mathematics skills in primary school. In this context, language skills turn out to be relevant for learning mathematics for students with non-German as well as those with German family language (Paetsch et al., 2016). Therefore, not only learners with non-German family language, but *all* learners with lower language skills should be considered in studies investigating learning mathematics.

The influence of language skills on mathematics performance does not necessarily imply that language skills are the cause of this relation. It is plausible that confounding variables, such as general cognitive abilities or socio-economic status, are involved. Investigations on general cognitive abilities[4] as a confounding variable showed that language skills explain the influence of language skills on mathematics performance beyond such general cognitive abilities (Heinze et al., 2007; Ufer et al., 2013). While the influence of general cognitive abilities on mathematics skills declines in the course of primary school years, the influence of language skills remains constant or even gains importance (Mücke, 2007). Alternatively, socio-economic status has been discussed as an explanation for differences in mathematics performance (e.g., Ufer & Bochnik, 2020). Students' socio-economic status can refer to their families' economic resources (e.g., supporting the students' learning by arranging a tutor) or their families' cultural resources (e.g., in the form of books). For the PISA study in 2015, Awisati and González-Sancho (2016) reported lower language skills for learners from families with lower socio-economic status. Furthermore, a similar relation between socio-economic status and mathematics performance was found (Baumert & Schümer, 2001; Ehmke & Jude, 2010). Several longitudinal studies in primary and secondary schools confirmed that differences in language skills predict differences in learning mathematics significantly, while the contribution of socio-economic status disappears when controlling for language skills (e.g., Ehmke, Hohensee, Siegle, & Prenzel, 2006; Ufer et al., 2013). Socio-economic status explains differences in mathematics performance at the end of the first school year, but the relationship

[4] According to Cattell (1963), *general cognitive abilities* can be subdivided into two components: fluid intelligence and crystallized intelligence. While fluid intelligence helps to solve new, unfamiliar problems, crystallized intelligence involves recalling knowledge acquired through prior learning.

between learning mathematics and socio-economic status decreases in longitudinal analyses (Ufer et al., 2013). Still, socio-economic status should be included as a background variable when investigating the role of language skills in learning mathematics (Ufer & Bochnik, 2020; Wilhelm, 2016). In summary, the observed relation between language and mathematics skills seems to go beyond the explanation by general cognitive abilities and socio-economic status. This emphasizes the importance of investigating the role of language skills when learning mathematics.

To understand how language skills influence learning mathematics, it is important to consider potential mechanisms, which could explain the relation between language and mathematics skills. In the literature (e.g., Bochnik, 2017), different explanations have been named, for instance using language in test situations, classroom discourse, and during thinking processes. These explanations and the respective state of research will be outlined in the following.

2.1.2 Mechanisms explaining the relation between language and mathematics skills

2.1.2.1 Linguistic demands in (written) test situations

One reason for language-related differences in mathematics performance could be the specific linguistic demands of mostly written test situations (Haag, Heppt, Stanat, Kuhl, & Pant, 2013; Martiniello, 2008; Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006). Lower language skills may lead to difficulties in understanding test items. However, it has been shown that targeted linguistic simplification of mathematical test items, for example by simplifying vocabulary or grammar, is not specifically effective for learners with lower language skills but equally effective for *all* learners (Abedi, Courtney, Leon, Kao, & Azzam, 2006; Kieffer, Lesaux, Rivera, & Francis, 2009). At the same time, test items with high linguistic demands seem to cause lower performance for all learners regardless of their language skills (e.g., Plath & Leiss, 2018). Due to these findings, it is unlikely that the observed language-related differences in mathematics performance solely originate from linguistic demands in test situations. Therefore, in order to explain language-related differences in mathematics performance, other mechanisms need to be considered. Instead of assessing mathematics performance, the next two mechanisms relate to learning mathematics.

2.1.2.2 Communicative use of language

The influence of language could also be explained through the *communicative use of language* in classroom discourse (Civil, 2008). Even if mathematics is often perceived as a subject that is not mainly based on language due to the intensive use of symbols, the meaning of specific symbols and representations still has to be linguistically negotiated in class (e.g., Steinbring, 1998). As Schütte (2009) illustrated in his dissertation using interaction analyses, this negotiation process is usually implicit, so following mathematics classroom discourse is often challenging. Qualitative studies from the U.S. context indicate that learners with lower language skills sometimes find it difficult to understand their teachers' utterances linguistically (Civil, 2008). Opportunities to interact with other learners may also be limited (Moschkovich, 2007). In a longitudinal study by Bochnik (2017), the mathematics skills of learners, who reported being able to follow classroom discourse, increased stronger in comparison to the mathematics skills of learners, who struggled in this matter. These findings provide initial evidence that language interaction in the classroom is an essential mediator between language skills and mathematics performance.

2.1.2.3 Cognitive use of language

Another explanation aims at the *cognitive use of language*. Language can serve as a tool for constructing and organizing mathematical knowledge (Maier & Schweiger, 1999; Sfard, 2008). Lower language skills can make it difficult for students to grasp mathematical concepts cognitively. Without adequate language skills, cognitive representations can only be built incompletely and further cognitive operations, such as transfer processes or problem solving, are only possible to a limited extent (Kempert et al., 2019). Using language cognitively plays a special role in more advanced mathematics that involve high-level cognition, such as word-problem solving (Peng et al., 2020). In order to infer mathematical structures from the verbal description of a situation, learners need to be familiar with mathematical concepts that match the described situation. However, acquiring such mathematical concepts can be limited, if lower language skills inhibit processes of constructing mathematical knowledge (Schlager, 2020).

In summary, language-related differences in mathematics performance can be attributed to linguistic demands in test situations and to using language during learning mathematics. However, Haag, Heppt, Roppelt, and Stanat (2015) rate the effects of linguistic simplification of test items as “small” in comparison to the effects of family language, socio-economic status, and language skills. They assume that the disadvantages of students with lower language skills in test situations rather originate from their learning processes. Such assumptions are endorsed by findings of a meta-analysis by Peng et al. (2020), which provide empirical evidence that both the cognitive and the communicative use of language play a role in the relation between language and mathematics. Thus, this work focuses on the students’ use of language during learning mathematics and ways of supporting students during this process.

To ensure apposite support, it seems helpful to understand, which language learners need when learning mathematics. Therefore, the next chapter gives an overview of general and subject-specific language needed for communicating about or constructing knowledge on mathematical concepts.

2.1.3 Forms of language needed for learning mathematics

In every school subject, students need general (subject-independent) and subject-specific language skills (Ufer & Bochnik, 2020). General language skills do not only refer to informal everyday language, but also to a more differentiated version of language. These two types of language have been referred to as *everyday registers* (also known as Basic Interpersonal Communication Skills, BICS; Cummins, 2008) and *academic language registers* (also known as Cognitive Academic Language Proficiency, CALP; Cummins, 2008). The term “register” targets the idea that language is used functionally and adapted to the requirements of different situations (Halliday, 1978; Meyer & Tiedemann, 2017): Depending on the context in which language is used (e.g., communication with friends vs. communication in the classroom), and depending on the respective goals, different registers and therefore different lexical, grammatical, and textual features are used. Since one goal of instruction is to make complex issues accessible for learners, the language that is used to communicate about those complex issues often draws upon more complex features (e.g., compound words, passive voice) than those that are used in everyday registers (e.g., Meyer & Prediger, 2012). Academic language registers facilitate systematizing, structuring, and expressing such complex issues precisely (Bochnik, 2017). While technical terms (e.g., sum, digit) are usually defined and discussed explicitly in the classroom, academic language registers often remain implicit (Schütte, 2009). In this matter, students may strongly benefit from learning opportunities outside school. This could be a disadvantage for students from families with a family language different from the language of instruction and with lower socio-economic status, since they may receive less learning opportunities to encounter and develop academic language registers (e.g., Heppt, Stanat, Dragon, Berendes, & Weinert, 2014).

Besides general language skills, also subject-specific language skills are discussed in literature to play a role in learning mathematics (e.g., Prediger & Wessel, 2013). Subject-specific language refers to the language that is used to communicate about a certain content area (Roelcke, 2010). For communicating about mathematical content, subject-specific language helps to express information precisely and completely (Meyer & Tiedemann, 2017). It has been investigated, if subject-specific language skills play a role in learning mathematics beyond general language skills (Wessel & Erath, 2018). Although subject-specific language skills overlap with general language skills and mathematics skills, they can be conceptualized and measured independently from both measures (Ufer & Bochnik, 2020). Within the dissertation project of Bochnik (2017), instruments to measure subject-specific language skills have been developed. These instruments focus on subject-specific vocabulary as well as text comprehension. Indeed, subject-specific language skills explain differences in learning mathematics beyond general language skills, and also beyond socio-economic status and general cognitive abilities (Bochnik, 2017).

To support students with developing subject-specific language, it is crucial to identify which language is used to describe a certain mathematical concept. For example, in the context of subtraction, words and phrases such as “to take away”, “to subtract”, or “to become less” should be accessible. Number decomposition could be described with “to be made up of”, “tens”, or “ones”. Niederhaus, Pöhler, and Prediger (2015) have collected such subject-specific *linguistic means*[5] for the content area of percentage calculation in secondary school. Schindler, Moser-Opitz, Cadonau-Bieler, and Ritterfeld (2019) have investigated linguistic means for basic operations (addition, subtraction, multiplication, and division) and geometry in primary school. However, since relevant linguistic means vary depending on the respective content area, more topics need to be approached. The systematic analysis of Schindler et al. (2019) already indicated that learners are often not familiar with terms that occur in everyday language as well as in the school context (e.g., “square”), and terms with different meanings in the everyday and the school context (e.g., “difference”). Systematically collecting linguistic means for different topics as in the mentioned study could be the first step to create a more systematic understanding, how teachers can develop techniques to access linguistic means for various topics.

Moreover, the influence of subject-specific language skills still needs to be investigated in more detail. It remains unclear, to which extent subject-specific language skills influence certain content areas and facets of mathematics skills. It is conceivable that subject-specific language skills may impose different requirements during word-problem solving than during equation solving. In contrast to equations, word problems are based on verbal descriptions of arithmetic situations, which can influence the solution process strongly. Subject-specific as well as general language skills seem particularly important for the solution of word problems: Reconstructing arithmetic concepts from verbal descriptions poses a major challenge for learners, especially for learners with lower language skills (Daroczy et al., 2015). Because of their specific demands, word problems were chosen as the focus of this dissertation project. In the following chapters, these specific demands and the role of language in solving word problems will be outlined.

2.2 Solving word problems

2.2.1 Additive one-step word problems – A definition

There is a long tradition of national and international research on word-problem solving (Daroczy et al., 2015; Kintsch & Greeno, 1985; Stern, 1998; Vicente, Orrantia, & Verschaffel, 2008). Typical word problems contain verbally described mathematical problems, which can

[5] The term *linguistic means* summarizes single words and phrases that are typically used to describe a certain topic or content area.

be solved by applying mathematical operations (Verschaffel et al., 2000). Contrary to context-free arithmetic tasks (e.g., “How much is 5 plus 7?”), such word problems describe mathematical operations with 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 that can be described by a mathematical concept (e.g., increasing a set) and their verbal description. Existing works often report on additive one-step word problems in the traditional sense (Breidenbach, 1969; Franke & Ruwisch, 2010), which are also the focus of this work. These problems can be solved with a single arithmetic operation (addition or subtraction[6]) and do not contain irrelevant information. The advantage of this format is that researchers can control task features systematically and focus on certain word problem types. However, the results obtained in this way do not allow direct conclusions about other sub-processes (e.g., interpreting and validating the outcome, Verschaffel et al., 2020) involved in solving more complex real-world problems (Kaiser, 2017). Typical examples for additive one-step word problems can be found in Chapter 2.4.2 (Fig. 2).

2.2.2 Theories on word-problem solving

Common theories on word-problem solving (e.g., Blum & Leiß, 2007; Kintsch & Greeno, 1985) assume that learners construct models when solving word problems. This idealized modeling process can be described as “transformational” (Czocher, 2018): Learners transform a real-world problem into a mathematical problem and back again. In the context of additive one-step word problems, two different types of models play a key role in the solution process (Kintsch & Greeno, 1985): 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 individually. The situation model is the learner’s internal, mental presentation of the given situation (Czocher, 2018). When learners describe this situation model with mathematical concepts, they construct a mathematical model, which is referred to as a “representation that is expressed externally and mathematically” (Czocher,

[6] Following the literature (e.g., Verschaffel et al., 2020), the term “additive” is used in this dissertation for situations and situation structures as a generic term for situations that can be modeled with addition or subtraction. For mathematical models and structures, however, this work differentiates between “additive models or structures” (containing addition as an operation, for example also subtraction as indirect addition) and “subtractive models or structures” (containing subtraction as an operation).

2018, p. 139). This representation can be, for example, an equation, a schematic drawing, or a graph. Students then process the mathematical model to obtain a certain numerical result. According to the modeling cycle by Blum and Leiß (2007), interpreting and validating the achieved results are important steps when solving complex, authentic real-world word problems. *Validating* the individually constructed models refers to the activity of “examining whether (or the extent to which) [they] are adequate” (Czocher, 2018, p. 137). Czocher (2018) suggests to extend established theories (e.g., Blum & Leiß, 2007), which view validating as a check at the end of the modeling process, by the idea of validating as an ongoing activity. Through this activity, learners may rate their initial models as incorrect and revise them subsequently (Zawojewski, 2013). However, in the context of additive one-step word problems that mainly focus on the emphasis of mathematical concepts, such simplified situations do rather not encourage such steps (Kaiser, 2017). Thus, these steps are backgrounded in this work.

The described theories are based on a cognitive perspective on modeling (Kaiser, 2017), which concentrates on the analysis of students’ construction of models instead of pedagogical or curricular goals (Czocher, 2018). This perspective of the *students’* individual solution processes can be contrasted with the structures that the *author* intended when creating a word problem. Based on this, a new framework including both perspectives – the learner’s and the author’s perspective – has been developed within this dissertation project (see Fig. 1, framework on structure levels of word problems).

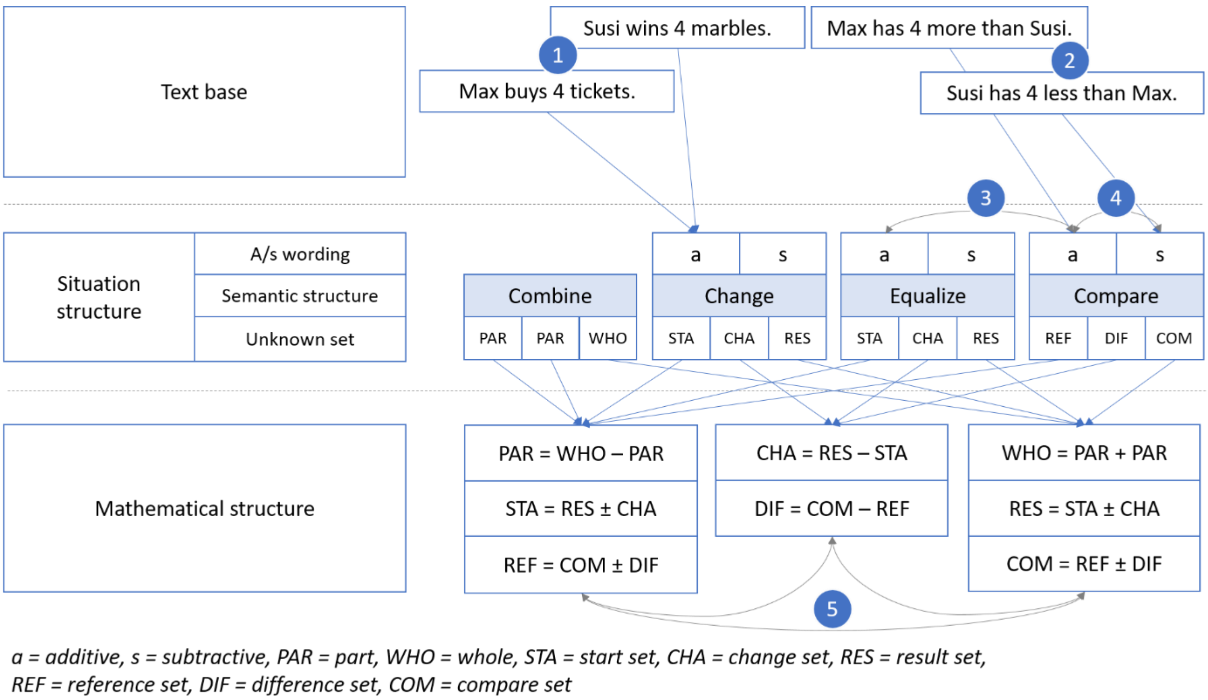

Fig. 1: Framework on structure levels of word problems

The framework addresses the two perspectives in the following way: On the one hand, it can depict the structures, which the author intended (for an example, see Fig. 4); on the other hand, it can depict structures that learners reconstructed individually as part of their individual situation model and mathematical model (for an example, see Fig. 5). These structures can either be assigned to the level of *situation structures* or *mathematical structures.* Learners decode situation structures and mathematical structures as accurately as possible during word-problem solving by forming an individual situation model and an individual mathematical model from the *text base*. Accordingly, the framework distinguishes two sub-processes, during which learners reconstruct these intended structures more or less completely when solving a word problem:

(1) The first sub-process relates to the connection of *text base* and *situation structure* (see Fig. 1). Authors can verbalize the problem's situation structure in different ways in form of a text base. Different text bases (Fig. 1, example ①) can express different situations but the same situation structure, and different verbalizations of the same situation (Fig. 1, example ②) can highlight different features of the situation structure. At first, learners decode the words and sentences of this text base and integrate this information into an initial situation model (Kintsch & Greeno, 1985). In the best case, this initial situation model contains the features of the situation structure, which were realized in the text base (see Fig. 1; *situation structure* level). According to models of reading comprehension (e.g., Kintsch, 1998), learners do not only reconstruct the content of the text base, but also add inferences to their situation model in some cases. For example, learners can enrich their situation model with features of the situation structure (e.g., visualized in Fig. 1, example ③ and ④; see also below). During the reading process, students may omit information they consider unimportant for the solution and infer information from their own knowledge that is not directly displayed in the text but necessary for processing (Stern, 1998). Thus, the situation structure, which was decoded by learners at the beginning of the solution process, may initially correspond to the situation structure that is described in the text base. However, learners can further differentiate and enrich their situation model with inferences in the further course.

In the framework, two substantially different characteristics of the text base are assumed to influence what is represented in students' (initial) situation models. First, the text base can emphasize different *features of the situation structure* (see the three sublevels of the situation structure in Fig. 1). For example, the fact that Max has four objects (e.g., marbles, cookies) more than Susi (see Fig. 1, ③) 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 case. Consequently, learners could *reinterpret* situations on the comparison of sets as situations, in which the difference between the two compared sets is equalized by an action, and integrate this new perspective into their individual situation model. Alternatively, the text base could also mention instead that Susi has four objects less than Max (see Fig. 1, ④). Depending on which features of the situation structure are realized in the text base, learners may construct different individual situation models according to this framework.

Second, authors can describe the *same* features of the situation structure with different linguistic means (text base in Fig. 1). On the one hand, the description of different situations requires different linguistic means. For example, verbalizing the comparison of two sets requires relational terms such as “more”, “larger” or “higher”, whereas describing the increase of a set requires words such as “to get”, “to win”, or “to buy”. On the other hand, authors can make use of various linguistic registers to describe a situation, such as everyday registers, academic language registers, or subject-specific registers (Prediger & Wessel, 2013, see also Chapter 2.1.3).

(2) The second sub-process relates to the connection of *situation structure* and *mathematical structure*. The situation structure intended by the word problem author is usually associated with one or more intended mathematical structures. Students transfer their constructed situation model into a mathematical model by describing their situation model with mathematical concepts (Kintsch & Greeno, 1985). At best, the mathematical model generated in this way reflects the intended mathematical structures correctly (see Fig. 1). On the one hand, selecting a particular mathematical structure as a mathematical model is based on the learners’ situation model. On the other hand, selecting a mathematical structure depends on the learners’ conceptual knowledge: Learners need conceptual knowledge how they can describe situation structures with 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, Siegler, & Alibali, 2001). This knowledge about possible meanings of certain mathematical concepts is also described with the term "Grundvorstellungen" in German-language mathematics education (Blum & Leiß, 2005; vom Hofe, 1995). 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" (or "Grundvorstellungen") must be individually available and also activated, so that learners can mathematize situation models. For example, learners can represent situation structures that refer to situations on the comparison of sets, or so-

called compare schemata (Riley, Greeno, & Heller, 1983; Schipper, 2009), in their individual situation model and transform their model into a mathematical model.

In summary, constructing a situation model plays a central role in word-problem solving. This construction process is assumed important in theoretical models such as the text comprehension model by Kintsch and Van Dijk (1978), which emphasizes the importance of reducing and organizing information from the text base, and also supported by empirical evidence (e.g., Leiss, Schukajlow, Blum, Messner, & Pekrun, 2010; Stern & Lehrndorfer, 1992; Thevenot, Devidal, Barrouillet, & Fayol, 2007). 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 feature generating difficulty 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. Thus, constructing adequate, comprehensive situation models may facilitate the solution process. Despite the (idealized) separation of the two sub-processes, the literature emphasizes that linguistic, situational, and mathematical knowledge are not activated separately during the entire solution procedure, but interact with each other (Stern, 1998).

It is confirmed by various studies that students' processing of additive one-step word problems diverges strongly. Beyond environmental factors, such as teachers or scoring criteria, Daroczy et al. (2015) summarize two main factors influencing the students' performance in their theoretical process model on word-problem solving. Parallel to the two perspectives on word problems described above (student vs. author side), the students' solution process is influenced by individual features of the students ("individual attributes" in Daroczy et al., 2015) and by the inherent task features of the given word problem ("stimulus attributes" in Daroczy et al., 2015). In the following two chapters, both factors and their detailed influence on word-problem solving will be outlined.

2.3 Individual features influencing students' performance on word problems

There are several features on the student side, which influence students' performance on additive one-step word problems. For example, students' affects can influence learners' performance in solving word problems (Verschaffel, Depaepe, & Van Dooren, 2015). Such affects can relate to the students' emotions (e.g., feeling motivated or frustrated), their attitudes (e.g., being (dis)interested in word-problem solving), and their beliefs (McLeod, 1992). Their beliefs can be, for instance, related to mathematics (e.g., "Every word problem has a solution.") or their self-efficacy (e.g., "I am able to solve word problems.") (Verschaffel et al., 2015). Besides affective factors, also general individual features such as the influence of social background, domain-general abilities, language skills, and subject-specific skills and knowledge have been discussed in literature (e.g., Muth, 1984; Verschaffel et al., 2015; Vilenius-Tuohimaa, Aunola, & Nurmi, 2008). The following subchapters give an overview of findings on these general individual features.

2.3.1 Social background

As described in Chapter 2.1.1, more attention has been given to the role of students' *socio-economic status* since the results of the first PISA study in 2000. Connected with the human capital theory by Bourdieu (1983), socio-economic status is characterized by indicators relating to economic (e.g., financial means) and cultural capital (e.g., books). These indicators are interconnected since economic capital can be used to increase cultural capital. For the assessment of learners' socio-economic status, researchers have often used the "books-at-home-index", which gives information about learners' economic and (objectified) cultural capital by gathering data on their family-owned books (Paulus, 2009). In Germany, the differences in performance caused by socio-economic status turned out to be particularly large compared to other participating countries (Ehmke & Jude, 2010). Such differences could not only be found for mathematics skills in general, but also for word-problem solving in particular (Coley, 2002).

2.3.2 Domain-general abilities

In the context of word-problem solving, also domain-general abilities are discussed. For example, learning to solve word problems successfully depends on *fluid intelligence* (Renkl & Stern, 1994), which may help with solving new, unfamiliar problems (Cattell, 1963). Although learners may have already encountered the underlying mathematical problem before, and although a direct solution may be possible, decoding and understanding the given text base usually poses a challenge, which is new every time (Renkl & Stern, 1994). High fluid intelligence may support students with handling such new challenges. To measure fluid intelligence, tasks on inductive reasoning are used as an indicator in many intelligence tests (e.g., the Culture Fair Test "CFT 1-R", Weiß & Osterland, 2013). Inductive reasoning refers to the ability to recognize patterns and regularities from individual examples (Lenhard & Lenhard, 2011). This ability is also assumed to help with recognizing schemata in word problems (see Chapter 2.2.2).

Furthermore, the importance of *working memory* during word-problem solving is emphasized strongly. According to cognitive load theory (Paas, Van Gog, & Sweller, 2010), there is only a certain amount of space for cognitive processes because of the limited capacity of the individual's working memory. Due to additional processes, such as reading the text base or constructing mental models, word-problem solving requires higher working memory capacity than solely conducting mathematical operations (e.g., Kintsch & Greeno, 1985; Peng & Lin, 2019). Studies confirm that, as a consequence, students with a higher working memory capacity achieve better performance in solving word problems than others (e.g., Wang, Fuchs, & Fuchs, 2016). If learners encounter problems that are yet too demanding, skipping the construction of models and applying superficial solution strategies instead may be the only option for them (Páchová & Vondrová, 2021). Fung and Swanson (2017) examined the role of working memory by investigating whether the effects of working memory on word-problem solving were completely mediated by reading skills, arithmetic skills, and fluid intelligence. Indeed, they identified a significant direct path for the storage component of working memory.

A further domain-general ability that has been discussed more recently to influence students' performance on word-problem solving is their *inhibitory control*, which is defined as "the ability to ignore salient but unhelpful stimuli and responses" (Verschaffel et al., 2020, p. 9). Learners may fail to inhibit stimuli that originate from the author's choice of words or numbers. For example, certain key words (e.g., more, less) may be associated with certain operations (e.g., addition, subtraction), which may or may not be the correct solution path when processing a word problem (e.g., Páchová & Vondrová, 2021). In order to identify the correct solution, learners need to inhibit these stimuli and concentrate on the construction of an adequate, comprehensive situation model. Similarly, if students do not inhibit misleading stimuli, they may apply superficial solution strategies instead of engaging intensively with the situation displayed in a word problem. The mechanisms related to task features as stimuli (e.g., chosen language, number material) will be outlined in more detail in Chapter 2.4.

2.3.3 Language skills

As described in Chapter 2.1.2, *language skills* have been shown to influence learning mathematics, since learners need these skills to construct mathematical knowledge and to participate in classroom discourse. In the context of word-problem solving, language skills are particularly required in the form of reading skills because 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 content areas. 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 facets of reading skills, which are crucial to decode the text base and construct an accurate situation model from this text base (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 (Beal, Adams, & Cohen, 2010; Muth, 1984; Vilenius-Tuohimaa et al., 2008). Students with higher reading comprehension skills seem to create more accurate situation models and mathematical models (Leiss et al., 2010). The dependence on reading comprehension skills can also be explained in connection with working memory: Based on cognitive load theory, students with lower reading comprehension skills may require more cognitive resources for decoding the text base and consequently have less capacity left for activating mathematical concepts and finding an adequate mathematical operation to describe the problem (Barbu & Beal, 2010). This increased demand on working memory may also apply, when learners with lower language skills use language cognitively to restructure their situation model during word-problem solving, for example to correct inconsistencies in their situation model (Greeno, 1980; Van den Broek, Young, Tzeng, & Linderholm, 1999).

2.3.4 Subject-specific skills and knowledge

Besides the general features described in the last sections, students also need certain subject-specific skills and knowledge for word-problem solving. For the solution of additive one-step word problems, *basic arithmetic skills and knowledge* in the context of addition and subtraction is necessary to set up an equation and calculate the result. This does not only refer to purely technical skills, but also to students' understanding of number concepts. It is crucial to understand numbers as a composition of other numbers (part-whole relationship), and addition and subtraction as complementary operations (Renkl & Stern, 1994). For instance, dealing flexibly with such complementary operations may help with solving tasks, in which one number on the left side of the equal sign is unknown (e.g., "3 + __ = 5" or "__ – 2 = 3"). Researchers assume that learners develop understanding of relations between numbers, such as part-whole relationships or differences between numbers, at later stages of number concept acquisition (e.g., Krajewski & Schneider, 2009). Schneider, Küspert, and Krajewski (2013) allocate this process of understanding relations between numbers to the first two years of primary school. Empirical findings confirm that students with higher basic arithmetic skills and knowledge achieve a better performance in word-problem solving, even when controlling for language skills (e.g., Bjork & Bowyer-Crane, 2013 for grade 2; Muth, 1984 for grade 6). Although performing calculations is not the central purpose of solving word problems, this process is still decisive for obtaining the correct result.

Learners also need conceptual knowledge to describe their situation model with mathematical concepts ("schemata", see Chapter 2.2.2). This requires that learners have access to such mathematical concepts and the corresponding knowledge, which mathematical concepts may fit a situation. For example, when solving a word problem involving subtraction, understanding this operation as a decrease of a set may not always be sufficient. In certain situations, it may be helpful to view subtraction as a way to describe a difference between two sets (Wessel, 2015). A lack of conceptual knowledge may result in inappropriate problem representation (Morales et al., 1985). Morales et al. (1985) compared third graders' with fifth/sixth graders' conceptual knowledge on schemata, which are necessary for solving additive one-step word problems, and found that older students did not only solve the word problems more accurately, but also showed a higher conceptual knowledge in this matter. Not only age, but also language skills can play a role when constructing conceptual knowledge on schemata. Based on the SOKKE study, Heinze et al. (2007) hypothesize that language skills rather influence developing conceptual knowledge and mental representations than procedures such as calculating. If learners cannot use language as a tool to construct knowledge in prior learning processes, they may not be able to access conceptual knowledge during word-problem solving (Wilhelm, 2016).

2.4 Task features influencing the difficulty of word problems

Not only individual features of the students influence their performance on additive one-step word problems, but also task features of the word problem. Following the presented framework on structure levels (Fig. 1), the features of a word problem can contribute to its difficulty on three levels: the linguistic complexity of the text base, the realized features of the situation structure, and the arithmetic complexity of the mathematical structure. While some features are independent from others, which means that they can be manipulated independently, other features are interlinked and therefore not seperable (Daroczy, Meurers, Heller, Wolska, & Nürk, 2020). The contribution of these three levels to the difficulty of additive one-step word problems varies strongly and will be outlined in the following.

2.4.1 Linguistic complexity of the text base

When students decode the information given in the text base in order to construct a situation model (see Chapter 2.2.2), their understanding is not only influenced by individual skills, but also by the demands of the given text base. A high linguistic complexity of the text base requires higher working memory capacity (Daroczy et al., 2015), which may cause comprehension obstacles and influence the learners' construction of a situation model (Barbu & Beal, 2010; Plath & Leiss, 2018). As far as additive one-step word problems are concerned, linguistic complexity needs to be delineated from discourse-interactional complexity, which refers to the interaction of participants in discourse (Daroczy et al., 2015). For the context of this work, discourse-interactional complexity is not applicable, since the considered word problems are not comprehensive enough to include interaction. Therefore, this chapter focuses on the linguistic complexity of word problems, which can affect the text base on a word level, sentence level, and overall text level.

Word problems typically make use of linguistic means that can be allocated to academic language registers and subject-specific registers. On the word level, certain features, such as unfamiliar vocabulary, ambiguous vocabulary, and the proportion of complex or long words were found, mainly in the context of English language learners, to increase a word problem's difficulty (Daroczy et al., 2015; Martiniello, 2008).

On the sentence level, studies identified the influence of syntactical features, such as noun phrase length, number of prepositional phrases, using passive voice, and complex clause structure, on task difficulty (e.g., Martiniello, 2008). Classical versions of additive one-step problems, however, primarily consist of main clauses, short noun phrases, few prepositional phrases, and active instead of passive voice due to their linguistically simplified and unified nature (e.g., word problems in the style of Stern, 1998). Still, some types of word problems constitute an exception (e.g., problems on the equalization of sets use conditional clauses: "*If* Susi eats two cookies, *then* she has as many cookies as Max has.").

On the overall text level, the difficulty of a task depends on text length (Haag et al., 2013), which is rather uniform among additive one-step word problems. Including irrelevant information also has been found to decrease students' solution rates (Muth, 1992; Wang et al., 2016); however, studies on additive one-step word problems usually do not include information, which is not necessary for the solution. There are still features on the overall text level that play a role in solving additive one-step word problems. For example, placing the question before the text seems to increase students' performance on word problems (Thevenot et al., 2007).

In summary, many features of linguistic complexity may apply to more complex problems (e.g., real-world problems, Chapter 2.2.1), but only to a limited extent to additive one-step word problems. Linguistic complexity can be increased artificially (e.g., by using nominalization), but this is not the focus of this work. On the contrary, the word problems considered in work on additive one-step word problems are usually worded as simple as possible (following the example of Stern, 1998), resulting in unified vocabulary and sentence structure, and the elimination of irrelevant information.

However, linguistic complexity is not only caused by the linguistic features described above, which can all be manipulated independently from the situation structure or the mathematical structure of a word problem. Such complexity can also come from the mathematical concepts described in a word problem, since verbalizing complex mathematical concepts requires demanding and specific language (Snow & Uccelli, 2009). Here, linguistic features are interlinked with the situation structure and the mathematical structure. For example, describing a comparison of sets requires complex language such as expressing relations with relational terms (e.g., "more than"). In this sense, linguistic complexity caused by the realization of mathematical concepts still plays a role when solving additive one-step word problems that have been worded as simple as possible. In the following chapter, the influence of such mathematical concepts and other features of the situation structure on a word problem's difficulty will be outlined.

2.4.2 Features of the situation structure

In the eighties, researchers initiated the classification of word problems according to their situation structure (e.g., Cummins et al., 1988; Nesher, Greeno, & Riley, 1982). This resulted in different word problem types, which can be distinguished by three features of a word problem's situation structure: *semantic structure*, *additive or subtractive wording*, and *unknown set* (see Fig. 1 for an overview of the features). Empirical studies on the difficulty of word problems have confirmed repeatedly that students' solution rates vary strongly depending on the described features (e.g., Stern, 1998; Verschaffel & De Corte, 1997). The following sections provide an overview of the different features on the level of the situation structure and their influence on a word problem's situational difficulty.

Semantic structure: The same mathematical structure (e.g., an additive operation such as 5 + 4 = 9) can describe different real-world phenomena (Fig. 2). 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).

	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. 2: Semantic structures describing the same mathematical structure

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 equalized. While *dynamic* word problems (change, equalize) describe actions, combine and compare problems refer to *static* situations (Carpenter, Hiebert, & Moser, 1981; Riley et al., 1983). Taking into account the relationship between the involved sets, it is also possible to distinguish semantic structures entailing a part-whole relationship (change, combine) from semantic structures that involve two disjoint sets (compare, equalize) (Radatz, Schipper, Ebeling, & Dröge, 1996).

When constructing a mathematical model, being able to access such semantic structures by activating schemata (see Chapter 2.2.2) is discussed to essentially influence students' performance when solving a word problem. 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., Cummins et al., 1988; Riley & Greeno, 1988; Stern, 1992). There are several factors that are discussed to explain the particular difficulty of compare problems theoretically. For example, 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 the relation between quantities. Current models on number concept acquisition also allocate such semantic structures in later phases of development (under the term "relational number concept"; for an overview see Fritz, Ehlert, & Leutner, 2018). Moreover, identifying the compared sets and understanding the syntactic structure of the sentence at the same time (Schleppegrell, 2007) is linguistically demanding. In compare problems, the relation is either expressed in a statement ("Susi has three marbles more than Max.") or in the question ("How many marbles does Susi have more than Max?"). According to Fuson, Carroll, and Landis (1996), it is vital to derive from a relational statement, *which* quantity is more or less and *how big* the difference between the two quantities is. 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 that first graders investigated in a study achieved relatively high solution rates of 96%.

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, Fig. 2) 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", Fig. 2) or removing a quantity (subtractive wording, e.g., "to give away", "to sell”). Again, combine problems take a special role, since the a/s wording cannot be varied here.

By determining a word problem’s text base and therefore the a/s wording, the reader can only access a snippet of the situation structure (e.g., “Max has 4 marbles more than Susi”). Through this determination, the a/s wording automatically emphasizes a certain perspective on the situation, while it tones other perspectives down. Thinking of alternative descriptions, which contain an inversed a/s wording but still describe the same situation (e.g., “Susi has 4 marbles less than Max”), can be seen as an inference process that exceeds the information given in the text base. The respective feature remains hidden at first and can be added during the reconstruction of the situation structure. Therefore, the a/s wording is not located at the text base level, but at the situation structure level (see Fig. 1). Recognizing the equivalence of such statements with inversed a/s wording seems to be an essential challenge when dealing with difficult word problems (Stern, 1993). The direct influence of a/s wording on a word problem’s difficulty, however, has not yet been reported systematically.

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* (see Fig. 1; e.g., Stern, 1993). Their equivalents in dynamic situations are *start set*, *change set*, and *result set*. For combine problems, only the *whole* set and its *parts* (subsets) are distinguished. 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 (Riley & Greeno, 1988; Stern, 1992; Van Lieshout & Xenidou-Dervou, 2020). This may be connected with 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[7] (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., $15 – 8 = x$).

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). As described in the prior section, the unknown set determines the directly applicable mathematical operation (see also Fig. 1). 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.

[7] Even if the directly applicable mathematical structure for a given situation model is subtractive, students may transform it into an indirect addition (Fig. 1, ⑤) (Torbeyns, De Smedt, Stassens, Ghesquière, & Verschaffel, 2009).

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). Using a/s wording as a surface indicator for the required mathematical operation only proves successful for consistent, but not for inconsistent word problems. The application of such *key word strategies* seems to depend on students' inhibitory control (Lubin et al., 2016; Verschaffel et al., 2020, see also Chapter 2.3.2): Students are more successful in solving inconsistent compare problems, if they are able to inhibit the stimulus of a key word (e.g., add if more, subtract if less). Lubin et al. (2016) found this effect even for experts in mathematics (e.g., adults); however, this group seems to be more efficient at inhibiting the misleading strategy than non-experts. If students do not inhibit the application of a key word strategy, they may skip the construction of a situation model and base their decision which mathematical operation is adequate on the a/s wording. The findings on the consistency of word problems are supported by eye tracking studies, which have observed that some learners mainly focus on key words and deduce the mathematical operation directly from this operation (Hegarty, Mayer, & Green, 1992). However, the solution of inconsistent word problems requires a deep analysis of the situation and therefore the construction of a sound situation model (Scheibling-Sève, Pasquinelli, & Sander, 2020). Stern (1993) and Mekhmandarov, Meron, and Peled (1996) reported from their studies that retelling inconsistent compare problems was difficult for many first graders. In contrast, retelling consistent compare problems was mastered by the majority of students. This suggests that it is not only the application of key word strategies that leads to different solution rates, but also problems with grasping the situation described in inconsistent compare problems. The effect of consistency is mostly discussed and investigated in the context of compare problems, but may also be applicable to dynamic problems, such as equalize or change problems (as in Daroczy et al., 2020). Although most studies on the consistency of word problems focused on primary school children, Verschaffel (1994) could still observe this pattern 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 the early school years.

2.4.3 Arithmetic complexity of the mathematical structure

On the level of the mathematical structure, a word problem's difficulty may be influenced by the number material used in the presented situation (Daroczy et al., 2020). For example, embedding *carry operations* (operations that include a change of the tens digit; e.g., 8 + 7 = 15) can make it more demanding to conduct an additive operation (Deschuyteneer, De Rammelaere, & Fias, 2005; Nuerk, Moeller, Klein, Willmes, & Fischer, 2015), since they require higher working memory capacity and thus involve a higher cognitive load (Daroczy et al., 2015). The same mechanism applies to subtractive operations, when *regrouping operations* (a decade from the minuend must be "opened" to make ten ones, in case the units of minuend and subtrahend cannot be subtracted) are necessary (Fuson et al., 1997). These features can be manipulated independently from the text base and the situation structure, which means that the arithmetic complexity can be adapted as desired. It has been observed that students need more response time when solving tasks with carry/regrouping conditions (Artemenko, Pixner, Moeller, & Nuerk, 2018), which indicates higher arithmetic complexity. Also, calculating with single-digit numbers is easier for most students than calculating with multi-digit numbers (Nuerk et al., 2015). However, the given number material may not only influence learners in implementing their calculation process, but also in constructing a mathematical model. Processing higher numbers may require more cognitive resources than processing smaller numbers, which may make it more difficult for learners to construct a mathematical model (Stern, 1998). The number of consecutive calculation steps also plays a role: Multi-step word problems are proven to be more demanding than one-step word problems (Muth, 1992; Quintero, 1983). Since this work focuses on one-step word problems, this factor is not elaborated further.

As far as the operation itself (addition vs. subtraction) is concerned, students were reported to achieve higher solution rates in solving word problems with an additive mathematical structure than solving those with a subtractive mathematical structure, resulting in shorter response times (Daroczy et al., 2020). However, these findings related to adults and only change problems. Thus, a systematic analysis in this matter still needs to be conducted for primary school children and for all types of additive one-step word problems.

As stated in Chapter 2.3.4, most students may have developed basic arithmetic skills and knowledge in the context of additive operations with numbers up to 100 in the course of the first two school years. Especially the first half of the second school year focuses strongly on the consolidation of such basic arithmetic skills and knowledge in additive contexts to build a sound foundation for multiplication and division. Therefore, it could be assumed that the basic arithmetic skills and knowledge needed for solving additive one-step word problems are available in most cases. Still, many studies report difficulties of learners with additive one-step word problems (e.g., on the level on the situation structure, see Chapter 2.4.2). Prior research has shown that the way an arithmetic problem is presented influences the difficulty for learners: In the school context, the same problem presented in numerical format (e.g., $3 + 5 = 8$) is solved 10 to 30% less frequently, if it is embedded in a word problem (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1980). This implies that fostering students' arithmetic skills is not sufficient when supporting students with word-problem solving (Stern, 1998). Rather, support should aim at fostering the students' understanding of arithmetic situations by constructing sound situation models.

2.4.4 General features of word problems

Besides the task features that can be assigned to one of the three levels of the framework (see Fig. 1), there are also overarching features, which may influence a word problem's difficulty. For example, the context, in which a word problem is embedded, may play a role. Davis-Dorsey, Ross, and Morrison (1991) reported that personalization (e.g., using students' own names in the word problems or familiar objects) led to higher solution rates for second graders. The authors hypothesized that such personalization could increase students' motivation and reduce demands on students' working memory. However, since these effects only appeared for word problems that were reworded additionally (using more explicit wording), it is not possible to clearly trace back these effects to personalization (Davis-Dorsey et al., 1991). Many other studies did not vary context features (involved subjects, objects, and numbers) independently from the features of the situation structure. This calls for a systematic variation of context features to separate their effect on task difficulty.

Another feature that may influence a word problem's difficulty is the form of representation. Word problems are verbal descriptions of an arithmetic situation, which are typically presented in text form. The text can be accompanied by nonlinguistic visual representations (Martiniello, 2009). According to a study by Martiniello (2009), adding schematic representations that emphasize mathematical relationships mitigated the impact of linguistic complexity on task difficulty for the participating fourth graders, in particular for those, who were not proficient in the language of instruction. A study in Northern Cyprus gained similar findings for fifth graders, in particular for word problems with a context, which was unfamiliar to the students (Cankoy & Özder, 2011).

In summary, there are various findings on task features influencing a word problem's difficulty. Many studies emphasize that learners' successful solution of word problems is dependent on their understanding of the given situation and their constructed situation model (see also Chapter 2.2.2). Stern and Lehrndorfer (1992) found that compare problems proved less difficult, when students were told a story compatible to the qualitative comparison depicted in the given situation. Therefore, they assume that it cannot be the linguistic complexity of such compare problems *alone* that is responsible for their particular difficulty. As stated before, the arithmetic complexity is rather secondary as a factor influencing the difficulty of a simple additive one-step word problem. Rather, studies report relatively consistent findings that emphasize the influence of the features of the situation structure, which were realized in the text base (Cummins et al., 1988; De Corte & Verschaffel, 1987; Stern, 1998): The task difficulty varies strongly depending on which semantic structure is described, which a/s wording is realized, and which set is unknown. Especially the last two features suggest that it is mostly of relevance, which features of the situation structure are realized, and less, how they are described verbally in form of the text base. Accordingly, this work assumes that it is not so much 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"). Consequently, it seems promising to focus on developing ideas how to enhance learners' understanding of arithmetic situations.

Prior interventions have already collected valuable information, which measures may support learners in this context. Students seem to benefit from a focus on relationships, patterns, and structures of arithmetic situations (Hasemann & Stern, 2002; Huang, Zhang, Chang, & Kimmins, 2019), but also from interventions that enhance understanding subject-specific language related to arithmetic situations (e.g., relational terms, Schumacher & Fuchs, 2012). As suggested implicitly in prior research (Greeno, 1980; Stern, 1993), learners could also be supported to enrich their situation model with inferences, so that this model can be mathematized more easily. For this, it may be helpful to develop flexibility in dealing with arithmetic situations occurring in additive one-step word problems. This approach and the associated construct will be outlined in the following chapter.

2.5 Flexibility in dealing with arithmetic situations

2.5.1 Defining the construct

Choosing an adequate mathematical operation is contingent on which features of a situation structure are included in the students' situation models (see Fig. 1): Depending on which features of the originally intended situation were reconstructed, it may be more or less straightforward to construct a mathematical model, and also the type of the mathematical model may vary. To overcome the reported barriers of students to mathematize their individual situation model, learners may benefit from conceptual knowledge, which may facilitate choosing an adequate mathematical operation (see Chapter 2.2.2). Based on the idea that conceptual knowledge comprises knowledge about interrelations (Rittle-Johnson et al., 2001), it may be helpful for learners to organize their conceptual knowledge by emphasizing connections between different word problem types with different situation structures.

This dissertation draws on these ideas and contributes a newly developed construct to the field. In this sense, *flexibility in dealing with arithmetic situations* can be defined as the skill to enrich situation models of additive one-step word problems with further features of the situation structure (a/s wording, semantic structures, unknown sets; see Chapter 2.2.2). 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 if a description fits the verbally presented situation or not. If learners struggle during word-problem solving, they may use flexibility to spontaneously restructure their knowledge (as described in the theory of "cognitive flexibility", Spiro, Feltovich, Jacobson, & Coulson, 1991).

Learners with a highly developed flexibility may generate descriptions of the given situations that (1) describe the situation accurately, and (2) include the activation of further features of the situation structure (e.g., an alternative description of the a/s wording). Such enriched situation models may support learners with constructing mathematical models. A low flexibility in dealing with arithmetic situations may be a reason for difficulties with certain word problem types. If students could be supported in developing the described flexibility, this may be an approach to help students with difficult word problem types. Practical observations (Fromme, Wartha, & Benz, 2011) as well as theoretical considerations (Greeno, 1980; Stern, 1993) endorse approaches based on the idea of flexibility. In the following sections, two promising, theoretically motivated approaches will be outlined.

2.5.2 Strategies for developing flexibility

Already in the early eighties, research suggested introducing strategies to reinterpret and enrich situation models with further information, so that they can be mathematized more easily (Fuson et al., 1996; Greeno, 1980; Stern, 1993). In this work, strategies are understood as cognitive procedures that have a heuristic value when solving a certain type of problem. Based on the mentioned suggestions, two strategies that may lead to the pursued flexibility can be deduced (Fig. 3): The *Inversion Strategy* and the *Dynamization Strategy*.

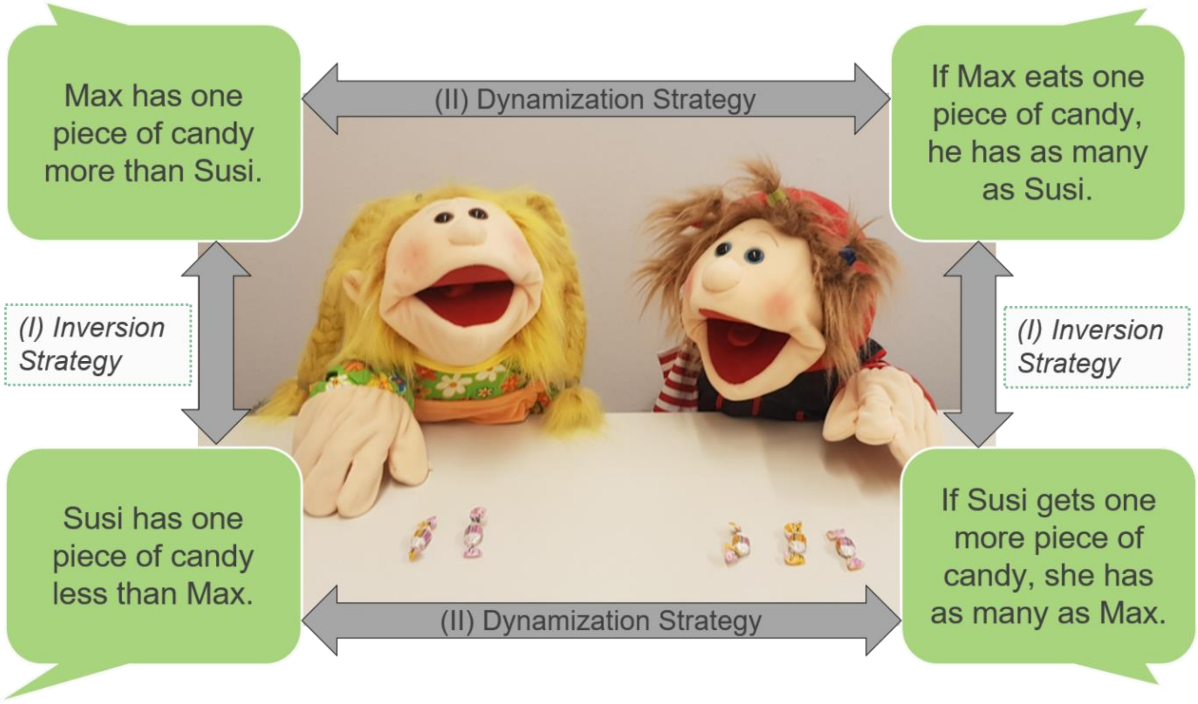

Fig. 3: Examples for Inversion Strategy and Dynamization Strategy

Inversion Strategy: Changing the perspective on mathematical relations. Stern (1993) investigated, if learners applied key word strategies during word-problem solving by asking them to retell compare problems. The results showed that learners could retell compare problems with unknown compare set (consistent) better than those with unknown reference set (inconsistent). Stern (1993) concluded from these findings that key word strategies cannot be the only reason for different solution rates regarding a problem's consistency. Instead of being the source of the problem, key word strategies may be an indicator that learners cannot make use of alternative strategies, since their conceptual knowledge on additive situation structures is not sufficient. In turn, Stern (1993) and other researchers (Fuson et al., 1996; Verschaffel, 1994) stress the role of understanding relational statements. 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 the linguistically symmetrical 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 ones with an unknown compare set (Fig. 1, ④; Fig. 3). To apply the Inversion Strategy, students need to reverse the subject and the object, and invert the a/s wording. This may pose high demand on the learners' working memory (Verschaffel, 1994). Although this has only been discussed for the case of compare problems, it is plausible that students may also benefit from understanding the symmetry of actions in the context of change and equalize problems.

Dynamization Strategy: Changing the semantic structure. Another suggestion aims at reinterpreting difficult semantic structures as easier accessible structures. Greeno (1980) proposed to help learners with 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 reported difficulties of students when solving compare problems, this idea could be transferred to a similar strategy (Fig. 1, ③, Fig. 3). Reinterpreting static compare problems into dynamic equalize problems seems particularly obvious here, since these situations both contain disjoint sets, but equalize problems are classified at a lower difficulty level than compare problems in models of word problem difficulty (Nesher et al., 1982; Radatz, 1983). 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. Students could *dynamize* compare problems by reinterpreting them as equalize problems, since dynamic equalizing may be easier to represent than a static comparison (see Fig. 3).

Dynamization and Inversion Strategies both aim at enriching the learners' situation model with further inferences while reconstructing the situation structure, so that the situation model can be mathematized more easily. The two strategies rely on conceptual knowledge, which is necessary to solve word problems (Morales et al., 1985). It helps learners to focus on relevant features of the situation structure and to add this information to their situation model (Rittle-Johnson et al., 2001). Developing flexibility may be one way to achieve this: If learners struggle with solving a difficult word problem, the inclusion of other perspectives on the situation may help them to create a more accurate and elaborate situation model.

For example, the described strategies could help learners with reinterpreting difficult compare problems with unknown reference set (see example in Fig. 4) as empirically easier problems (e.g., equalize problems with unknown result set). In the framework on structure levels (Fig. 1), the application of both strategies could be visualized the following way.

When learners work on the solution of a difficult compare problem with unknown reference set (see Fig. 4), they may encounter difficulties to find an adequate mathematical structure that fits their individual situation model. Instead of relying on key word strategies (Chapter 2.4.2), students could resort to strategies to describe the given situation flexibly.

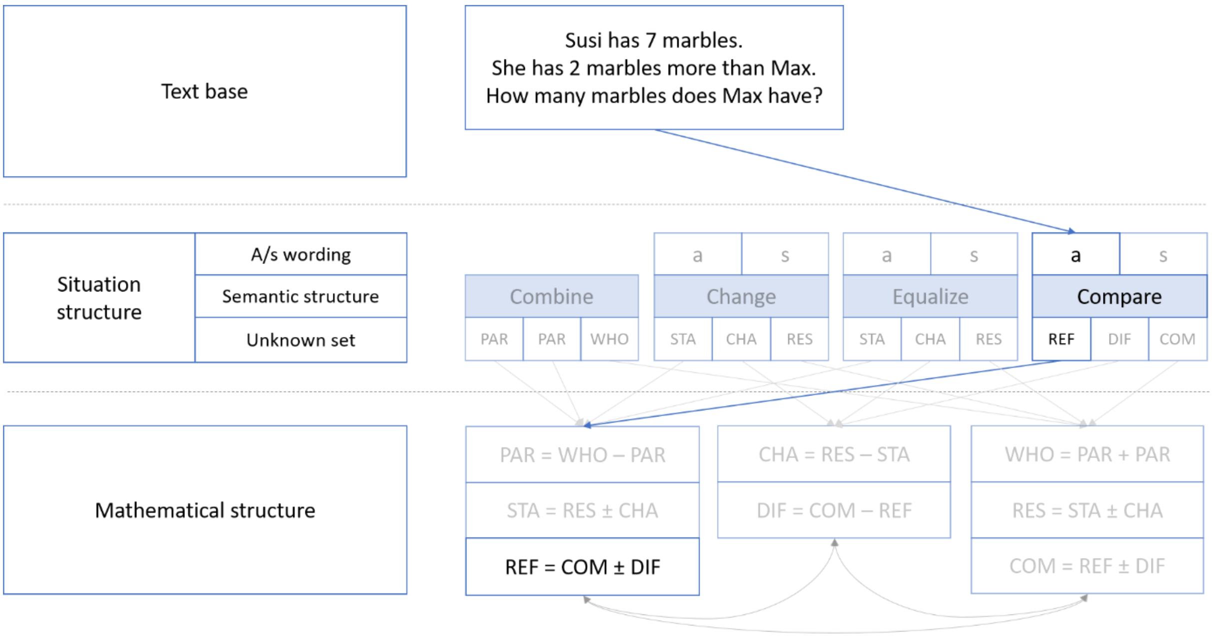

Fig. 4: The visualization of a compare problem with unknown reference set *without* including further features of the situation structure

If students describe the situation flexibly and activate further features of the situation, students may reinterpret the situation as an equalization of sets (by applying the Dynamization Strategy) and/or describe the situation with a subtractive wording (by applying the Inversion Strategy) (see Fig. 5). Their reinterpretation could be carried out verbally (by actively describing the new features and perspectives with words) or mentally (by using language cognitively). By adding the information "If Susi gives two marbles away, she has as many marbles as Max has", the compare problem with unknown reference set could also be interpreted as an equalize problem with unknown result set.

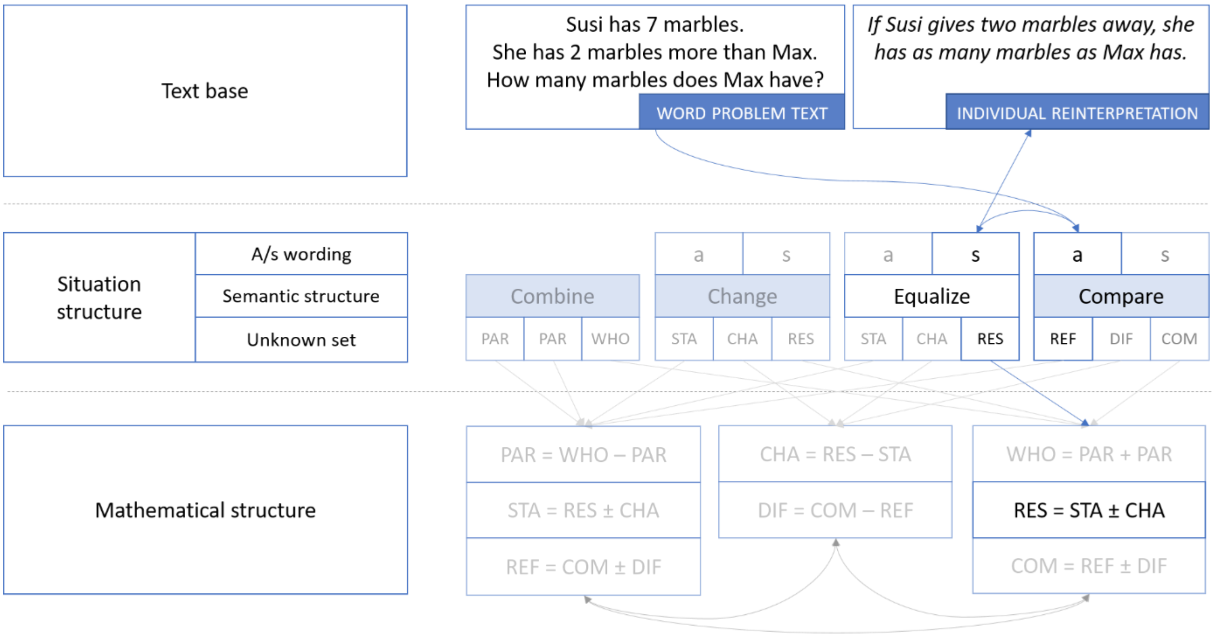

Fig. 5: The visualization of a compare problem with unknown reference set with the inclusion of further features of the situation structure (equalization through dynamization, subtractive wording through inversion). Learners may add a new description verbally or mentally (see cursive script on the text base level)

It is a central assumption of this work that this new focus may improve the conditions for learners to solve difficult word problems. Although strategies have already been proposed in literature, no corresponding interventions to test this assumption have been conducted. The following chapter proposes instructional approaches, which may be appropriate for supporting learners in developing flexibility in dealing with arithmetic situations by promoting the Inversion Strategy and the Dynamization Strategy.

2.5.3 Instructional approaches for developing flexibility

Traditional research on additive one-step word problems has strongly focused on identifying language-related features influencing their difficulty (e.g., Daroczy et al., 2015). Of course, the consequence cannot be to avoid these features in instruction. On the contrary, it is vital to find ways to support students in understanding also more difficult problem types. Given their findings that using language cognitively is crucial during word-problem solving in particular, Peng et al. (2020) suggest that instructions should facilitate students' reasoning and inference making. One way may be to foster flexibility in dealing with arithmetic situations and encourage learners to enrich their situation model by making inferences (see Chapter 2.5.1). If an increased flexibility indeed enhances the students' understanding of additive one-step word problems, it is desirable to find ways to support students in developing such flexibility. For this, learners may need to acquire conceptual knowledge on features of the situation structure, but also the language to communicate about them or to organize their knowledge cognitively. Therefore, instructional approaches to develop flexibility should enable learners to draw connections between mathematical concepts, and also to gain language skills for describing such conceptual connections. Such language skills should not be restricted to knowledge of individual word meanings, but also include “rich semantic networks with robust connections between the meanings of words associated by topic” (Currie & Cain, 2015, p. 59), which may facilitate making accurate inferences when constructing a situation model. This could be achieved by comparing different descriptions of the same, but structurally different situations, which may stimulate learners to make connections between different features of the situation structure and thus enrich their conceptual knowledge of the underlying arithmetic operations. To support the process of developing flexibility as described, two instructional approaches seem to be particularly promising: integrating a learning trajectory and enhancing language.

2.5.3.1 Integrating a learning trajectory

Integrating a learning trajectory as an instructional approach can serve as a guide for teachers when making instructional decisions (e.g., Wilson, Sztajn, Edgington, & Myers, 2015). Such learning trajectories try to predict probable learning paths of students (Simon, 1995). Of course, the students’ individual learning paths may differ from this hypothetical learning trajectory. A learning trajectory comprises a learning goal, specified learning activities, and a hypothetical learning process during these activities (Simon, 1995). In the context of word-problem solving, research has already demonstrated how a well-planned learning trajectory can enhance students’ understanding (Huang et al., 2019). For the learning goal of developing flexibility in dealing with arithmetic situations, no learning trajectory has been identified yet.

This dissertation suggests two dimensions, in which the hypothetical learning trajectory may progress: (1) the level of situational difficulty and (2) the level of flexibility. Following the literature review (Chapter 2.4), instructional approaches for developing the pursued flexibility should predict a hypothetical learning process that progresses from easier to more difficult word problem types (1). Learners may start with analyzing connections between different descriptions of easier arithmetic situations (e.g., easier in terms of the unknown set, or the form of representation), and then progress to working with empirically more challenging given situations (Huang et al., 2019; Riley & Greeno, 1988). Assuming that the pursued flexibility develops gradually, learning processes should also progress from initial stages of flexibility to more advanced forms (2). Flexibility may emerge as a receptive ability, which may be complemented by a productive component in later stages of development. At first, it may be easier for learners to engage with different *given* descriptions of situations (*receptive flexibility*). Actively describing a situation in various ways (*productive flexibility*) may then build on this receptive ability.

Corresponding learning activities should be aligned to this assumed learning trajectory. Since flexibility includes questioning the validity of interpretations (Warner et al., 2003), learners could be encouraged at first to verify, if given statements are correct. To analyze contrasts between different possibilities to describe an arithmetic situation, learners could investigate similarities and differences between such statements. This may draw connections between different ways of describing a situation, and also between different features of the situation structure. After building receptive flexibility, learners may practice to actively describe arithmetic situations comprehensively and adequately on their own. Based on these guidelines, developing a learning trajectory with the learning goal of developing flexibility and corresponding learning activities needs to be taken into account when choosing an instructional approach.

2.5.3.2 Enhancing language

Understanding and expressing descriptions of situations flexibly is tightly connected to language skills. Therefore, instructional approaches should consider the role of using language for word-problem solving in three different ways:

(1) Flexibility in dealing with arithmetic situations entails to transform or rephrase difficult problem types into easier problems, which learners can already approach successfully. This strategy, which has been proposed by other authors (Greeno, 1980; Stern, 1993), targets *using language cognitively when solving* word problems. Instead of simplifying difficult word problem types by rewording them beforehand (e.g., Vicente et al., 2008), instructional approaches could amplify students' language use by encouraging them to make inferences on the level of the situation structure (Erath, Ingram, Moschkovich, & Prediger, 2021; Schleppegrell, 2007).

(2) Students will only find this strategy helpful, if they are sensitive to the ways in which different descriptions of a situation are related to each other. It is assumed in this dissertation 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, it is expected that analyzing how language is used in different ways to describe arithmetic situations provides fruitful learning opportunities. This can be considered as an example of *using language cognitively when learning* about the features of such situation structures (Götze, 2019).

(3) Finally, to achieve exactly this analysis, instructional approaches should encourage students to use *communication* for learning processes (Moschkovich, 2015). Students could verbalize different descriptions of the same situation, reason why these descriptions fit the same situation, and explain and discuss structural similarities and differences between situations.

In summary, communication can be used to create learning opportunities for using language cognitively. During this process, the communicative use of language has a reinforcing effect on the cognitive use (Maier & Schweiger, 1999).

Snow and Uccelli (2009) state that dealing with complex mathematical concepts requires demanding and specific language. To support students in dealing with such demands, *enhancing language* is assumed to be a promising approach. This may provide learners with corresponding subject-specific linguistic means for describing situations flexibly (Pöhler & Prediger, 2015). Erath et al. (2021) summarized six design principles for designing materials and instruction that aim at enhancing language for learning mathematics. According to the first design principle, teachers should "(P1) … engage students in rich discourse practices". Instead of teaching only mathematical vocabulary, activities should support the development of conceptual knowledge, for example by explaining, arguing, or justifying a certain situation (Moschkovich, 2015). Teachers can support this process by providing learners with opportunities that encourage discourse in multiple modes (receptive and productive use of language, oral and written forms of language) (Erath et al., 2021). Instruction material should facilitate dealing with multiple modes and emphasize mathematical concepts. Learners can be supported by materialized scaffolds (e.g., language frames, visuals). For a successful implementation of this design principle, teachers may support learners' interactions and discussions continuously.

Another way of enhancing language is to "(P2) … establish various mathematics language routines" (Erath et al., 2021, p. 247). Through these routines, learners may assess their own and their peers' language production in "a structured but adaptable format" (Zwiers et al., 2017, p. 9). Teachers can provide feedback during such routines so that learners can revise and refine their language (Erath et al., 2021). An established example for a language routine in mathematics instruction is the format "Always, sometimes, never" (Swan, 2003), in which students decide, if a mathematical statement is always, sometimes, or never true.

Enhancing language also entails to "(P3) … connect language varieties and multimodal representations" (Erath et al., 2021, p. 247). This means learners should encounter mathematical concepts in the form of different registers (everyday registers, academic language registers, technical terms; see Chapter 2.1.3) and various representations (e.g., symbols, graphics, diagrams…). By making connections, learners can relate to aspects they are already familiar with. Erath et al. (2021) emphasize the importance of *connecting* language varieties and representations instead of changing between, or even progressing only linearly. To make connections during learning, learners could be encouraged to argue about connections between verbal descriptions and representations (Prediger & Wessel, 2013).

With respect to learners' family language, which may differ from the language of instruction, it is advised to "(P4) … include students' multilingual resources" (Erath et al., 2021, p. 247). Drawing on descriptions and explanations in another, potentially more familiar language allows to make use of additional resources (Erath et al., 2021).

A further approach to enhance language and familiarize learners with linguistic means relevant for a particular mathematical subject is to "(P5) … use macro-scaffolding to sequence and combine language and mathematics learning opportunities" (Erath et al., 2021, p. 247). Macro-scaffolding describes pre-organized support by the teacher considering students' different language skills (Hammond & Gibbons, 2005) and entails a sequencing of tasks that allows students to progress from accessible tasks 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 instance, visualizations or specific language support (Prediger & Pöhler, 2015). Making use of materials following the principles of scaffolding facilitates the combination of language and mathematics learning (Pöhler & Prediger, 2015).

Finally, Erath et al. (2021, p. 247) recommend to "(P6) … compare language pieces (form, function, etc.) to raise students' language awareness". Language awareness refers to being conscious of language on a meta level, for example by gaining explicit knowledge about language, or perceiving language consciously (García, 2017). In the context of enhancing language when learning mathematics, learners can compare and contrast pieces of subject-specific language to become sensitized to the subtleties of language (Erath et al., 2021). On the word level, they can draw connections between similar words or words belonging to the same word family. On the sentence level, similarly structured sentences can be compared and contrasted to emphasize syntactical features (Dröse & Prediger, 2019). On the overall discourse level, it is also possible to compare entire explanations or argumentations. Analyzing such contrasts relates to variation processes, which are connected with the variation theory (Kullberg, Kempe, & Marton, 2017; Pang, Bao, & Ki, 2017). This theory sees changing perspectives as the essence of learning, since it encourages learners to compare and contrast features of the studied object (Huang et al., 2019).

In the following passages, it is discussed, which role the listed design principles could play in the specific context of developing flexibility in dealing with arithmetic situations.

To develop conceptual knowledge during the analysis of different descriptions, it can be helpful to enhance rich discourse practices (P1, Erath et al., 2021). Explaining, justifying, and arguing about various descriptions of word problems may help with recognizing interrelations between features of the situation structure. For this, learners should encounter flexible descriptions targeting different manifestations of semantic structures, a/s wording, and unknown sets.

Material should address multiple modes, in which learners are encouraged to read and listen to flexible descriptions (and their explanation or justification) but also to formulate their own flexible descriptions (and their explanation or justification) orally or in written form. This may provide rich language input by their tutor and their peers along with rich language output by the students themselves.

In the context of developing the pursued flexibility, it may also be helpful to establish mathematics language routines (P2). When learners work with descriptions of a situation (produced by themselves, a peer, or in the form of pre-organized material), they could be encouraged to verify, if this description matches the given situation. Such verifying activities may enable self and peer assessment, and also the revision and refinement of given descriptions. Erath et al. (2021) indicate that language routines need to be structured and adaptable. This could be ensured well in the suggested format “verify, then refine or revise”, since teachers could decide to adapt the difficulty and complexity of given situations and descriptions depending on the learners’ abilities. For example, they could progress from descriptions of qualitative to quantitative comparisons, or they could progress from given situations on comparing two concrete sets in a picture to situations, which correspond to compare problems with an unknown reference set in text form (see also the hypothetical learning trajectory in Chapter 2.5.3.1).

Connecting language varieties and multimodal representations (P3) could also support students when developing flexibility (Erath et al., 2021). Besides the classical representation of word problems in text form, a situation can also be re-enacted with the students, portrayed in pictures, or visualized with manipulatives. Also purely verbal descriptions of a situation may be a form of representation. Connecting these multimodal representations with potential descriptions of a situation not only enhances the development of conceptual knowledge (Schleppegrell, 2007), but also emphasizes the connection between different options to describe an arithmetic situation. Concerning the connection of language varieties, it may be helpful for learners to interconnect everyday registers and academic language registers required for describing situations in word problems (Prediger & Wessel, 2013). To draw this connection, students and tutors may analyze connections between their language productions (which may rather tend to consist of everyday language) and the language in provided descriptions, which were designed with the aim to serve as a language model.

Describing arithmetic situations flexibly may pose a challenge for learners with lower language skills in particular. For multilingual learners with lower language skills, it may help to include other languages they speak that are different from the language of instruction (P4) (Erath et al., 2021). If learners struggle with describing a situation in the language of instruction, teachers may encourage them to have recourse to their mother tongue. Ideally, the teacher or peers are able to address the learner's attempt. However, this is only possible, if teachers and peers are familiar with this language.

To make use of macro-scaffolding, the linguistic means related to the two strategies should be sequenced to lead learners to a well-connected, rich vocabulary (P5). The Inversion Strategy and the Dynamization Strategy build on vocabulary on relational terms and action verbs. Comprehensive vocabulary on pairs of antonyms (e.g., bigger and smaller) or inverse actions (e.g., to give and to get) may be helpful for a precise, rich description of the given situation. In addition, the emphasized semantic structures are embedded in specific syntactical structures. Expressing equalization statements relies on action verbs and conditional clauses ("If …, then …"). Situations on the comparison of sets require formulating relational statements (e.g., "Susi has two marbles more than Max"). Language frames (e.g., on a sentence level) could be used as scaffolding material to draw attention to crucial parts of descriptions. At first, these sentence frames may provide a stronger support by asking only for single words or numbers. With the students' experience increasing, the support may be faded out by asking for longer phrases. In the end, the aim is that learners do not need the scaffolding material anymore to describe situations flexibly.

To raise language awareness, Erath et al. (2021) suggest to include the comparison of language pieces (P6). Comparing and contrasting descriptions of situations that follow the Inversion Strategy or the Dynamization Strategy could sensitize students to the linguistic means used in different descriptions and the subtleties of their interpretation. For example, relational statements have a syntactical symmetry (Stern, 1993), which can be highlighted through direct comparison. Fig. 6 gives an example how different relational statements could be contrasted systematically.

Susi	has three marbles	more	than	Max
Max	has three marbles	less	than	Susi
Susi	has three marbles	less	than	Max
Max	has three marbles	more	than	Susi

Fig. 6: Contrasting relational statements systematically (bordered columns emphasize the critical parts of relational statements that need to be varied to invert the statement)

These statements either describe the same situation (statements within the first resp. the second pair) or opposing situations (comparing the first pair of statements with the second pair of statements). Parts of the sentence that are critical for applying the Inversion Strategy (involved subjects, relational terms; see bordered columns in Fig. 6) may be asked for in sentence frames (macro-scaffolding, see P5). On the word level, it may help to analyze word families on relational terms and action verbs. In the German language, relational terms can be transformed into action verbs by adding a prefix (e.g., “kleiner”, “verkleinern”[8]). Drawing connections between such words may also help with connecting compare situations with equalize situations.

[8] This roughly translates as “smaller” and “make smaller”.

3 Overview of the dissertation project

3.1 Relevance of the topic

Word problems in the traditional sense (see Chapter 2.2.1) could be criticized for being unrealistic and irrelevant for students' life. However, solving traditional word problems in the mathematics classroom makes a unique contribution to learning mathematics. Stern (1998) summarizes two major goals of using word problems in primary school: (1) using word problems for learning modeling, and (2) using word problems for learning mathematical concepts.

(1) Additive one-step word problems can serve as an entry point for developing modeling skills. Contrary to real-world problems, which also emphasize other modeling steps, such as interpreting and validating the result, traditional word problems concentrate on constructing a situation model and a mathematical model (see Chapter 2.2.1). Due to this restriction, students can practice these crucial construction processes with a clearer focus (Verschaffel et al., 2000). In the context of inquiry-based learning, working with additive one-step word problems is not simply based on applying calculations (Franke & Ruwisch, 2010), but students are encouraged to make connections between different operations during modeling, for example, by manufacturing their own math stories (e.g., "Which math story fits to 5 + 6?").

(2) Because situations in word problems are reduced in their situational complexity, using word problems in mathematics instruction facilitates to emphasize mathematical concepts better than using real-world problems. Dealing with mathematical concepts in word problems prepares learners for subsequent, more demanding mathematical issues in secondary school. For example, understanding numbers not only as a tool to count or to define a quantity, but also as a tool to describe the relation between two numbers is important for advanced mathematics learning (Stern, 1994). Similar to difference sets, fractions describe a relation between two numbers. Indeed, when dealing with fractions in secondary school, it has been observed that students, who performed better in solving compare problems during the first primary school years, performed better than students, who achieved lower performance in solving compare problems (Stern, 1994). Dealing with word problems can also enhance understanding of part-whole relations (Resnick, 1989), which includes understanding addition and subtraction as inverse operations as well as understanding numbers as a composition of other numbers (e.g., "7" is also "5 + 2") (Renkl & Stern, 1994). This conceptual knowledge is the foundation for applying calculation strategies, solving algebraic equations, and understanding functions (Stern, 1994). In school, simple arithmetic equations with a blank before the equation sign (e.g., __ + 5 = 9) are often used to emphasize the relationship of addition and subtraction. However, students may develop stereotypical strategies to solve such arithmetic equations without building solid understanding, if such equations are not

connected with mental representations (Stern, 1994). Teachers may use manipulatives to build mental representations of arithmetic operations (e.g., as in Wartha & Schulz, 2011), but also embedding such equations in word problems may advance the students' conceptual knowledge of addition and subtraction (Stern, 1998). However, this only seems possible, if students do not rely on key word strategies and instead engage with the displayed situation by constructing a situation model.

The described goals illustrate why word problems should be an important element of modern mathematics education. Current classroom practice seems to deviate from this claim in some aspects, at least if textbooks are used as a reference point: While empirically easier types of word problems (e.g., combine and change problems, problems with unknown result set) are a regular feature of textbooks, more difficult types (e.g., compare problems) are rather scarce, as analyzed in a thesis with ten textbooks for Bavarian schools (Von Damnitz, 2020). However, learning opportunities in textbooks influence students' performance in mathematics, which could be confirmed for several content areas (e.g., Sievert, Van den Ham, & Heinze, 2021; Van den Ham & Heinze, 2018). This finding together with the reported difficulties on the level of the situation structure (Chapter 2.4.2) suggests that instructional approaches to help students overcome their difficulties demand more attention.

As described in Chapter 2.5.2, strategies for developing flexibility that can be integrated in instructional approaches have already been suggested (e.g., by Greeno, 1980; Stern, 1993), but such approaches have not been implemented yet. The dissertation project addresses this research gap and investigates the construct of flexibility in dealing with arithmetic situations as a promising approach to support students dealing with difficult additive one-step word problems, in particular inconsistent compare problems (see Chapter 2.4.2). Fostering learners' flexibility may help them to enrich their individual situation model with further features of the situation structure and then to mathematize their model (see Chapter 2.5.1). According to Schoenfeld (2007, p. 60), good problem solvers "have many ways to think about problems – alternative approaches if they get stuck." By enhancing conceptual knowledge through interconnecting features of the situation structure, this project gives students alternatives if they "get stuck" during word-problem solving. This is assumed to make key word strategies redundant and shift the focus on situational understanding. Using language for learning mathematics is particularly prominent in this context, since learners may not only need to acquire conceptual knowledge on features of the situation structure, but also the language to communicate about them or to organize their knowledge cognitively (see Chapter 2.5.3.2). Thus, language-sensitive instruction is an essential part of this approach. In the following chapter, the goals of the dissertation project will be outlined based on the literature review.

3.2 Goals of the dissertation project

3.2.1 Goal 1: Revisiting a long-standing field of research

The field of research on additive one-step word problems is relatively established in mathematics education. Emerging in the eighties, many studies and projects have tried to reconstruct students' processes during word-problem solving, and to identify and explain differences in difficulty and in individual performance between tasks (e.g., Carpenter et al., 1981; Kintsch & Greeno, 1985; Nesher et al., 1982). Subsequently, researchers have tried to manipulate task features of word problems (on the level of the text base, the situation structure, and the mathematical structure) to investigate the influence of this manipulation on a word problem's difficulty (e.g., Stern, 1993; Vicente, Orrantia, & Verschaffel, 2007). This is an important contribution to understand the mechanisms behind the effects of such task features, but it does not address ways to enhance students' understanding. Instead of avoiding difficult task features beforehand, research is necessary to identify adequate instructional approaches to help students *deal* with the reported difficulties. Several effective interventions on word problems already exist (e.g., Hasemann & Stern, 2002; Huang et al., 2019; Schumacher & Fuchs, 2012), but many build on data that were collected a long time ago. Mathematics instruction, and thus students' difficulties may have progressed since then. Therefore, it is necessary to (1.1) replicate prior findings on the difficulty of the identified word problem types (following the formulation of Stern, 1998), before an intervention study can be conceptualized and conducted. In case prior findings no longer mirror the current situation, this may influence the concept and design of prospective interventions.

Another sub-goal, which intended to help updating prior results, is to (1.2) systematize prior findings. Methodically, the reported studies usually did not vary the features of the situation structure independently from other features, such as the names of the subjects, the involved objects, or number material. Even if only small effects are expected here, investigating potential performance differences caused by these features is still pending.

3.2.2 Goal 2: Investigating the construct of flexibility in dealing with arithmetic situations

Replicating and systematizing prior results can establish a basis for investigating the newly proposed construct of flexibility in dealing with arithmetic situations. It is not clear yet, if some learners already acquire flexibility spontaneously[9] at least to some degree. An instrument to measure flexibility in dealing with arithmetic situations was successfully developed within a thesis (Weber, 2016), but it still needs to be investigated, if students with a high, in principle available flexibility also use this flexibility spontaneously during word-problem solving, when

[9] "Spontaneously" refers to the idea that learners may already make use of flexibility without explicit instruction within an intervention or similar support.

the task format encourages this spontaneous use. Therefore, it is necessary to (2.1) provide learners with situations that stimulate the use of flexibility and investigate if students use such flexibility spontaneously for solving word problems, even before an intervention is conducted.

Once this is clarified, it makes sense to conceptualize an intervention. The main goal of this dissertation project is to develop an approach to help students deal with difficult additive one-step word problems. Derived from suggestions by Stern (1993) and Greeno (1980), developing flexibility in dealing with arithmetic situations is assumed in this dissertation to be a promising approach to reach this goal. Since this flexibility is a newly developed construct, the intervention is conducted as a "feasibility study" (2.2) and strives to investigate, if flexibility in dealing with arithmetic situations can be enhanced *at all* with the suggested instructional approaches (Chapter 2.5.3) . If this is the case, it does not necessarily imply that learners can make use of this flexibility when solving difficult word problems. Therefore, it is also of interest, (2.3) if enhancing flexibility in dealing with arithmetic situations has a positive effect on word-problem solving.

Flexibility in dealing with arithmetic situations is assumed in this work to be tightly connected with language (see Chapter 2.5.3): When students deal with descriptions of arithmetic situations, this likely puts high requirements on their language skills, since using language cognitively can help with analyzing, connecting, and contrasting features of the situation structure (Maier & Schweiger, 1999; Sfard, 2008). Due to the reported importance of language skills for solving word problems (e.g., Peng et al., 2020), an intervention should support students with lower language skills in particular. However, it is unclear, how students with different language skills will respond to the intervention. There is a risk that lower language skills may limit the intervention's effect. Therefore, the analysis also aimed at investigating, (2.4) which role language plays when developing the pursued flexibility, and how language skills influence the intervention's effect.

Another sub-goal is to (2.5) gain detailed insights into the students' development of flexibility during the intervention. There may be certain key processes, which are crucial when developing flexibility. Comparing students' learning paths and examining, if the hypothetical learning trajectory fits to the students' actual development may help to understand the construct in more depth and adjust future programs accordingly.

The mentioned goals were approached within two studies (see Fig. 7 for an overview). A preliminary study intended to prepare the development of an intervention by revisiting the field of research (1.1, 1.2) and investigating, if students already make use of the pursued flexibility spontaneously (2.1). This preliminary study is reported in Chapter 4. Based on these results, an intervention program was developed. This intervention study is reported in Chapter 5. After outlining the design of the intervention study in Chapter 5.2, the goals 2.2, 2.3, and 2.4 to investigate the intervention’s effect will be pursued in the quantitative analyses of the study (Chapter 5.3). This is followed by a qualitative analysis to identify different key processes when developing flexibility and to explore the students’ development of flexibility during the program (2.5) (Chapter 5.4). After presenting and discussing the results of the two conducted studies, a general discussion will conclude with a summary, limitations, and implications for future research and teaching.

Goal 1 Revisiting a long-standing field of research	**1.1** Replicating prior findings on the difficulty of additive word problems	**Preliminary study** *Chapter 4*
	1.2 Systematizing prior findings	
Goal 2 Investigating the construct of flexibility in dealing with arithmetic situations	**2.1** Investigating the spontaneous use of flexibility	
	2.2 Investigating the intervention‘s effect on flexibility	**Intervention study: Quantitative analysis** *Chapter 5.3*
	2.3 Investigating the intervention‘s effect on word-problem solving	
	2.4 Investigating the effect of language skills on the intervention	
	2.5 Analyzing students‘ development during the intervention	**Intervention study: Qualitative analysis** *Chapter 5.4*

Fig. 7: Overview of the goals of the dissertation project and the conducted studies

4 Study on the spontaneous use of situation structures for solving word problems

4.1 Introduction

As outlined in Chapter 2.4.2, various studies reported differences in the difficulty of word problems depending on the features of the situation structure that are realized in a word problem. However, the majority of these findings date back to the eighties and nineties. Since then, mathematics education may have progressed significantly, for example in the context of teaching standards (e.g., the Standards for Mathematical Practice, National Governors Association Center for Best Practices, 2010), modeling (Greefrath, Kaiser, Blum, & Borromeo Ferri, 2013), and the role of word problems (Verschaffel et al., 2000). Furthermore, prior studies have often varied the features of the situation structure together with other features of a word problem. This makes it difficult to trace back differences in difficulty solely to the features of the situation structure. Consequently, it is necessary to validate and systematize the available findings.

Previous findings on differences in the difficulty of word problems with different situation structures (e.g., Riley & Greeno, 1988; Stern, 1998) suggest that learners do not make use of the pursued flexibility in dealing with arithmetic situations (see Chapter 2.5.1) completely spontaneously or without problems when solving word problems yet. Two major explanations may be plausible. On the one hand, it is possible that learners did not develop flexibility in dealing with arithmetic situations yet. On the other hand, it is also plausible that learners do not use this strategy in general, or in certain situations, although they have the necessary skills. For example, it would be possible that learners only use the strategy spontaneously, if they already constructed an alternative situation model and assigned it to a mathematical structure. In this case, solving successive word problems that are structurally similar could stimulate learners to use the strategy. In terms of the framework on structure levels in Fig. 1, *structurally similar* refers to the idea that learners can reinterpret one situation (e.g., a comparison of sets) as the other situation (e.g., an equalization of sets) by making inferences at the level of situation structures. Flexibility in dealing with arithmetic situations and using this skill strategically could hence manifest itself in the way that learners use structural similarities to relate the solution of a more difficult word problem back to an easier, preceding word problem. Using such strategies is intended to simplify more difficult word problem types (e.g., compare problems), especially when learners can enrich the situation structure of the previous, easier task with inferences, so that it matches the presented text base of the following, more difficult word problem. In this case, it is no longer necessary to construct a completely new situation model. In principle, it is sufficient to match the text of the second word problem with the already constructed situation model from the previous word problem by making inferences on the level of situation structures. In this way, learners may recognize that they could use the situation model from the previous task, and thus the identified mathematical structure, again. If students can indeed make use of structurally similar word problems, this would make an intervention study redundant and raise other questions, for example, how structurally similar word problems could be used during mathematics instruction.

Another reason, why learners may not use the pursued flexibility yet, could be so-called *socio-mathematical norms*. Such socio-mathematical norms are informal norms about how tasks should be approached in mathematics classrooms, which are socially constructed in classroom interaction (e.g., Yackel & Cobb, 1996). For example, learners may assume that mathematical solution approaches can and must always be generated only from the given word problem without considering previous word problems. Such a norm would also explain, why learners often apply superficial solution strategies during word-problem solving (e.g., Verschaffel, De Corte, & Lasure, 1994). If a short hint that learners could make use of similar structures was sufficient to significantly improve learner performance, this would speak for the second explanation – and question the need for an intervention study on developing the pursued flexibility.

The concept of flexibility in dealing with arithmetic situations (Chapter 2.5.1, Chapter 2.5.2) is promoted implicitly in the work of Greeno (1980) and Stern (1993). This dissertation project proposes to understand this flexibility as an individual ability construct, and to examine it for learner performance, feasibility of fostering this ability, and to what extent this flexibility is relevant for students' performance on word-problem solving. The preliminary study presented in Chapter 4 aims at capturing the current situation, if and how learners already make use of this flexibility when solving word problems.

4.2 Aims and research questions

To study if an intervention based on Greeno's and Stern's suggestions could be helpful, two main issues need to be resolved: The first aim of this preliminary study is to replicate and systematize prior results regarding task difficulty, which are fundamental for the intervention (see Chapter 3.2.1, goals 1.1 and 1.2).

Q1: Which of the task features *semantic structure*, a/s *wording,* and *unknown set* cause differences regarding the difficulty of additive one-step word problems?

Based on prior studies (e.g., Cummins et al., 1988; Stern, 1992), it was expected that compare problems were more difficult than equalize (H1.1) and change problems (H1.2). Solution rates should be higher for consistent word problems than for inconsistent word problems (H1.3). Since the direct influence of a/s wording on a word problem's difficulty has not yet been reported systematically, no explicit prior hypotheses were made. Moreover, existing studies have not varied context features of the word problems (involved subjects, objects, and numbers) independently from the mentioned task features. This study controls the variation of difficulty caused by these context features. Only minor differences in solution rates due to context features were expected (H1.4).

The second aim of this study was to investigate, if learners use the described strategy to reinterpret situation structures spontaneously (see Chapter 3.2.2, goal 2.1).

Q2: Do students use similar situation structures in subsequent word problems spontaneously to solve word problems? Does a hint on the structural similarity support learners with using the strategy?

Using this strategy successfully should cause higher solution rates for items for the second of two structurally similar, consecutive word problems (as compared to the first problem in the pair, H2.1). Stronger differences were expected, when compare situations occurred after a dynamic situation, than in the reverse sequence (H2.2). In the case that students have already gained the required knowledge but do not apply the strategy spontaneously, it was assumed that the effects in H2.1 and H2.2 would be more pronounced, if the learners received an explicit hint on the similarity of the situation structures (H2.3).

4.3 Method

To answer the research questions, paper-and-pencil based tests were used in a cross-sectional study with second graders from eight classrooms in three schools in Munich, Germany (N = 139). With 48% female subjects, the gender distribution was roughly balanced. At the time of the data collection, which was the middle of the school year, the students' average age was 7.56 years. The language of instruction was German. To gather information on the students' family language, self-reports by the students were collected. These indicate that 40% of the participants speak exclusively German at home with their family, 43% of the children speak German and another language, and 16% of the children only speak one or more languages at home, which are different from German (one missing value).

4.3.1 Material

Each student solved different additive one-step word problems, which were selected from a larger collection of task variations. To examine Q2, the word problems were arranged in pairs. Each word problem pair contained two structurally similar word problems with the same context (e.g., names of the subjects, involved objects), mathematical structure, a/s wording, and unknown set. The pairs differed only in their semantic structure. The word problems within each pair hence described structurally similar, but yet different situations. For example, a word problem pair could consist of task A: "Susi has 13 marbles, Max has 8 marbles. How many marbles does Susi have to put away, so that she has as many as Max has?” and task B: “Susi has 13 marbles, Max has 8 marbles. How many marbles does Max have less than Susi?” In this example, both tasks contain the same mathematical structure (13 – 8 = 5), the same context (Susi and Max as subjects, marbles as involved objects), the same type of unknown set (difference set), and the same a/s wording (subtractive). The only difference is the semantic structure: The first word problem deals with the equalization of sets, while the second word problem describes a similar comparison of sets. To generate all possible versions of word problem pairs, combinations of semantic structures (change and compare, equalize and compare), the a/s wording, and the unknown set were varied systematically. The linguistic complexity of the text base, which realized the particular situation structure, was controlled as far as possible by reducing the variation of vocabulary and syntax. These prototypical types of word problems were each embedded in twelve different situation contexts consisting of different persons, involved objects, and number material. In the end, a second set of word problems was generated by reversing the order of the two word problems within each pair.

This design was based on the hypothesis that learners could benefit from reinterpreting compare situations as change or equalize situations. These last two semantic structures turned out to be less difficult than the comparison of sets in prior studies (e.g., Stern, 1994). Also in models on number concept acquisition (Fritz et al., 2018; Krajewski & Schneider, 2009), understanding the underlying concepts of change and equalization are allocated to earlier stages of development than understanding relations between numerical quantities as in compare problems. Equalize problems may be particularly suitable for reinterpretation due to their structural similarity: Both equalize and compare problems contain disjoint sets. Combine problems were not considered, since they turned out difficult to combine with compare problems in a thesis (Weber, 2016).

4.3.2 Procedure

The data were collected with paper-and-pencil based tests within each classroom. Each student solved twelve randomly selected word problem pairs in a random sequence. In each questionnaire, the second and the seventh word problem pair were replaced by a distractor pair, which had dissimilar mathematical structures in the two tasks (adding instead of subtracting the two given numbers, and vice versa). This was intended to avoid that students solved only the first task and automatically transferred the answer to the second word problem without actually reading the text. Each page of the questionnaire showed one word problem. Students were instructed to not move backwards through the pages. This intended to prevent students from adjusting their answers retrospectively. To solve all twelve word problem pairs, students had 40 minutes of time. After half of the time, the students took a short break.

In half of the participating classrooms, students received an explicit hint, which aimed to encourage them to use similar structures for solving the following word problem. In these classrooms, the instructor explained beforehand that some tasks were quite similar and that some word problems described the same situation like in the task before, just with other words. Students were encouraged to transfer the equation from the prior task instead of starting a new calculation, if they recognized such situations: “I have a little hint for you. Some tasks are about the same thing as the task before, it is just described a little differently. In this case, you don’t need to calculate again, but you can write down the same solution as in the task before.” This experimental manipulation aimed at monitoring, if socio-mathematical norms or students’ beliefs about the expected problem-solving process were the reason, why students potentially did not make use of their abilities to deal flexibly with arithmetic situations.

4.3.3 Coding

Students’ solutions were coded in two different ways: The first option (correct result) classified the answer of a student as correct, if the numerical result was correct. The second option (correct operation) classified the answer of a student as correct, if at least the calculation path *or* the result was correct. In this context, all equations and calculation strategies, which are mathematically equivalent to the word problem’s mathematical structure, were classified as a correct operation. For example, this could be “bridging through ten” (e.g., $7 + 8 = 7 + 3 + 5 = 15$) (Thompson, 1999), changing the order of the addends, or solving subtraction problems by means of indirect addition.

4.3.4 Statistical analysis

For the inferential statistical analyses, generalized linear mixed models for dichotomous data with a logit link function were used (Bates, Mächler, Bolker, & Walker, 2014), which predict the correctness of an operation or a result for each task based on individual person features and task features. Dependencies between answers of the same person were taken into account by including a random intercept. Dependencies between students’ responses on word problems with the same context and responses by learners from the same classroom were modeled with respective random intercepts, if they explained a proportion of variance, which was identifiably different from zero. To examine main effects and interaction effects of the task features, likelihood ratio (LR) tests based on a chi-square statistic were used. This test compares the fit of the model *with* the respective effect with the fit of the model *without* the effect. To compare solution rates under different conditions, contrasts between the respective estimated marginal means were calculated. The reported regression coefficients can be interpreted as difference values on a log odds ratio scale similar to differences of item parameters in an IRT model. All calculations were executed in R with the packages *lme4* (Bates et al., 2014) and *emmeans* (Lenth, Singmann, Love, Buerkner, & Herve, 2018).

4.4 Results

4.4.1 Q1: Replication of prior results on task difficulty

To answer Q1, only the first task of each word problem pair was analyzed, excluding the distractor pairs. As expected (H1.4), the variation of the situation context explained only a small proportion of variance (less than 0.01%). In the following sections, the main effects of a word problem's semantic structure, unknown set, and a/s wording are presented.

Semantic structure: There were no significant differences between the semantic structures concerning the frequency of correct results (change: 77.1%, equalize: 71.0%, compare: 72.1%; LR test $\chi^2(2) = 4.06$; $p = 0.13$). However, students identified the correct operation significantly less frequently (B = 0.60; $p = 0.03$) in equalize problems than in change problems (change: 82.4%; equalize: 75.4%; compare; 76.3%; LR test $\chi^2(2) = 6.71$; $p = 0.027$). Concerning compare problems, there were no significant differences regarding the frequency of correct operations in comparison to change and equalize problems. These results did not confirm H1.1 and H1.2.

A/s wording: The main effects of a/s wording on the frequency of correct results (LR test $\chi^2(1) = 0.54$; $p = 0.46$) and the frequency of correct operations (LR test $\chi^2(1) = 2.70$; $p = 0.10$) were not significant. This finding fills a research gap, which has not been reported systematically yet.

Unknown set: The analysis revealed significant differences for the frequency of correct results (LR test $\chi^2(2) = 20.99$; $p < 0.001$) and correct operations (LR test $\chi^2(2) = 32.72$; $p < 0.001$) depending on the unknown set. Students gave the correct result significantly more often, if the result/compare set (78.1%; B = 0.51; $p < 0.001$), or the change/difference set (74.0%; B = 0.81; $p = 0.01$) were unknown, than if the start/reference set was unknown (66.8%). The difference between word problems with unknown result/compare set and those with unknown change/difference set was not significant (B = 0.30; $p = 0.22$). Similar effects occurred, when the identification of correct operations was analyzed: Students identified correct operations more frequently, if the result/compare set was unknown (84.3% correct operations), than when the change/difference set (76.3%; B = 0.78; $p < 0.001$) or the start/reference set (72.0%; B = 1.15; $p < 0.001$) were unknown. The difference between word problems with unknown start/reference set and those with unknown change/difference set was not significant (B = 0.37; $p = 0.14$). Hence, word problems with unknown start/reference set in particular proved to be more difficult than word problems with unknown result/compare set. This is in line with prior results by Stern (1993).

Interaction effects: Second, the interactions between the three main effects were analyzed. The results showed a significant interaction effect of unknown set and a/s wording for the frequency of correct results (LR test, $\chi^2(2) = 22.40$; $p < 0.001$) as well as the frequency of correct operations (LR test, $\chi^2(2) = 30.84$; $p < 0.001$). Furthermore, there was an interaction of semantic structure and a/s wording (LR test, $\chi^2(2) = 8.20$; $p = 0.017$). The three-way interaction was not significant for both performance measures (results: LR test, $\chi^2(4) = 2.75$; $p = 0.60$; operations: LR test, $\chi^2(4) = 2.61$; $p = 0.62$).

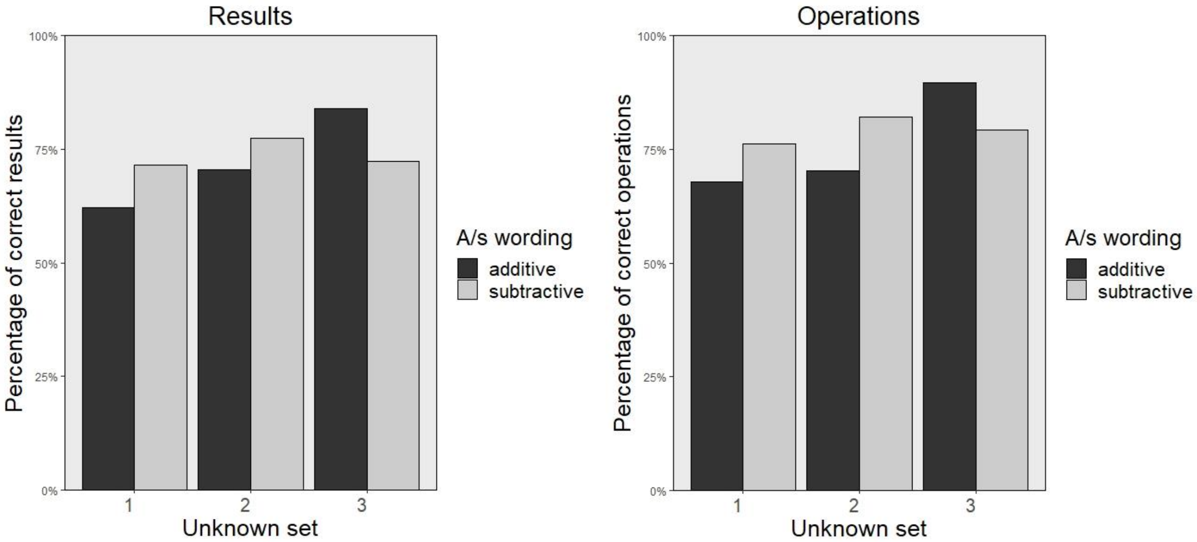

Fig. 8: Overview of the solution rates (left side: correct results, right side: correct operations) depending on the unknown set (1 = start/reference set, 2 = change/difference set, 3 = result/compare set) and a/s wording

Unknown set and a/s wording: Fig. 8 shows the percentage of correct results (Fig. 8, left side) and correct operations (Fig. 8, right side) depending on the unknown set and the a/s wording. As expected (H1.3), correct results and operations occurred more frequently for consistent word problems (unknown result/compare set, and unknown change/difference set with subtractive wording: 77.9% correct results; 83.6% correct operations) than in tasks on inconsistent word problems (68.0% correct results; 71.4% correct operations; planned contrast for results B = 0.76; $p < 0.001$; for operations B = 1.18; $p < 0.001$).

However, Fig. 8 also indicates that there are differences between additively and subtractively worded problems within each type of unknown set. If the start/reference set is unknown, the a/s wording and the directly applicable operation are inconsistent. If the result/compare set is unknown, the a/s wording and the directly applicable operation are consistent (see Chapter 2.4.2). Consequently, the consistency of a/s wording and the operation cannot be the reason for different solution rates. In both cases, students solved those word problems correctly more frequently that are described best by an additive mathematical structure. This was investigated with planned contrasts. For tasks that are described best by an addition[10] (additively worded problems with unknown result/compare set, subtractively worded problems with unknown start/reference set), students identified the correct result (77.9%; B = 0.54; $p < 0.001$) and the

[10] or an equivalent mathematical structure

correct operation (82.8%; B = 0.81; $p < 0.001$) significantly more often than for tasks that are described best by a subtraction[11] (70.6% correct results; 74.8% correct operations).

Semantic structure and a/s wording: Concerning the interaction of semantic structure and a/s wording, the contrasts between word problems with additive or subtractive wording showed no significant differences regarding the correct operation. Descriptively, students identified the correct operation less frequently in additively worded change (80.1%) and compare problems (73.0%) than in corresponding subtractively worded word problems (84.7% and 79.7%). For equalize problems, the reverse trend was identified (additive: 77.9%; subtractive: 72.9%).

4.4.2 Q2: Using situation structures as a spontaneous solution strategy

The aim of the second research question was to investigate, if learners already used similar situation structures for solving word problems spontaneously (without explicit instruction), and if a short hint encouraged this use. To this end, both tasks of each word problem pair were analyzed separately for the pairs with the combinations “change / compare” and “equalize / compare”. The main effects for the semantic structure indicated the same pattern as in the analyses regarding the first research question.

Task position: Based on the hypotheses, the effects of the task position in a word problem pair (first vs. second task in a pair) were analyzed. The main effects of task position were not significant for both coding options and both variations of word problem pairs (e.g., LR test for pairs of compare and change problems: correct results: $\chi^2(1) = 1.61$; $p = 0.20$, correct operation: $\chi^2(1) = 1.67$; $p = 0.28$). Consequently, the hypothesis that the processing of a structurally similar word problem beforehand supports students with solving the following task was not confirmed (H2.1).

Task position and semantic structure: Also, the interaction of task position and semantic structure was not significant in all cases (e.g., LR test for pairs of change and compare problems: correct results: $\chi^2(1) = 0.13$; $p = 0.72$, correct operation: $\chi^2(1) = 0.58$; $p = 0.81$). Thus, the hypothesis that solving change or equalize problems improves the solution rates of structurally similar, subsequent compare problems was not confirmed (H2.2).

Hint at similar structure: Finally, the effect of giving a hint on the similarity of word problems within each pair was included into the models. This main effect and its interaction with task position were neither significant for the combination of change and compare problems (e.g., LR test for correct results: hint $\chi^2(1) = 0.97$; $p = 0.32$; interaction $\chi^2(1) = 0.11$; $p = 0.73$) nor for the combination of equalize and compare problems. In addition, the interaction of hint and semantic structure and the three-way interaction of hint, task position, and semantic structure were not significant. Consequently, the hint showed no significant effect on using structurally similar word problems (H2.3).

Further analyses: The results regarding the first research question already suggest that it is not, as assumed, the semantic structure that primarily explains differences in the difficulty of word problems for the sample, but the a/s wording of the task combined with the unknown set. Therefore, further explorative investigations were conducted, whether interaction effects of a/s wording or unknown set with task position and hints could be detected for individual task types. Only pairs of equalize and compare problems showed an interaction effect of unknown set and task position (LR test: $\chi^2(2) = 6.18$; $p < 0.05$). Here, students identified the correct operation significantly more often in tasks with unknown change/difference set, when the task was second in the pair (81.3%) than when it was the first of the two tasks in a pair (76.1%; B = 0.38; $p < 0.05$). The present data did not provide evidence that this difference would have been particularly pronounced under certain circumstances, such as when the compare problem was second in the pair or when a hint at the structural similarity of the problems was given. Furthermore, this interaction effect of unknown set and task position was not significant when *correct results* were used instead of *correct operations* as the criterion variable (LR test: $\chi^2(2) = 1.91$; $p = 0.38$). No evidence was found that learners benefited more from equalize problems for solving compare problems or vice versa. The pattern of effects might suggest that students use the structural similarity between compare and equalize problems at least to some extent to build mathematical models.

4.5 Discussion

4.5.1 Q1: Replication of prior results on task difficulty

A first goal of this preliminary study was to investigate systematically, which of the factors reported in the literature influence the difficulty of word problems in the target population (goal 1.1 and 1.2). For this purpose, only the first tasks of each word problem pair were analyzed. Overall, the solution rates for different task types were more homogeneous and higher compared to prior studies (e.g., Stern, 1994 in grade 1). Context features, such as the subjects' names, the objects used in the task (marbles, flowers ...), or the specific number material, had little observable effect on task difficulty. Moreover, only minor differences were found according to the semantic structure of a word problem. This contrasts with prior findings by, for instance, Stern (1998) in grade 1, who identified compare problems as particularly difficult. It seems to be comparatively easy for the analyzed second graders to reconstruct the three semantic structures from the text base and to describe them with an adequate mathematical structure. One explanation could be that by the middle of the second grade, learners have already gained substantial experience with all semantic structures. Besides, they may benefit from a change in learning opportunities since the existing studies, for example, because different types of realistic tasks are treated more frequently in class (e.g., National Governors Association Center for Best Practices, 2010). Especially for compare problems, the latter explanation seems quite plausible, as quantitative comparison plays a central role in recent models of number concept acquisition (e.g., Krajewski & Schneider, 2009) and also in analyses of textbook quality (Sievert et al., 2021). The descriptive results of Fritz et al. (2018) further suggest that already about a quarter of students have knowledge on relations between quantities at the beginning of grade 2, which is likely to increase until the middle of the school year. The rather low solution rates for equalize problems compared to previous studies (Stern, 1994) could possibly be attributed to misunderstandings arising from the task wording. Students could interpret equalization as an act of equalizing *between* the involved persons ("How much does Max need to get *from Susi* so that they have the same amount of objects?"). In future studies, equalize problems should be formulated more precisely in order to exclude such an interpretation. Within this study, the answers of the participating students do not indicate such a misinterpretation.

Results from this study underline that the a/s wording combined with the unknown set is far more important for identifying a correct operation or result. Higher solution rates occurred when the a/s wording matched the directly applicable arithmetic operation. On the one hand, this may indicate that learners choose arithmetic operations solely based on key words occurring in the text base. On the other hand, a lacking understanding regarding the symmetry of relations, as addressed by Stern (1993), would explain learners' problems in constructing a situation model and thus, why they may have used key word strategies for compare problems. The fact that similar effects occur for dynamic semantic structures, such as change and equalization, might suggest that it was not only the understanding of numbers as relations (Fritz et al., 2018) that learners struggled with in this study. More likely, understanding the symmetry of relations (Stern, 1993) *and* actions in general seems to contribute strongly to students' success. If many learners already developed an understanding of the relational number concept at the time of this study, it would be an open question, whether understanding the symmetry of relations and actions emerges later in development, and if the development progresses simultaneously for static and dynamic semantic structures.

For the initially planned intervention program, this would mean that it should focus not primarily on the relationships between different semantic structures (e.g., compare and equalize problems). Rather, this program should shift the focus to the symmetry of relations and actions within the individual semantic structures. Descriptively, the study shows differences in identifying the correct operation between tasks with additive and subtractive wording, which varied between semantic structures. This also suggests to address symmetry equally for the three investigated semantic structures.

Furthermore, the analyses reveal another phenomenon that, as far as is known, has not been described in the literature before. Regardless of the a/s wording, the unknown set, and the semantic structure, learners in this study seem to solve word problems that require addition (or equivalent mathematical structures) as a mathematical model better than those that require subtraction (or equivalent mathematical structures). This was observed for the number of correct results. Initially, this could be explained by the different difficulty of the underlying additive or subtractive mathematical models, for example, by divergent counting directions in counting strategies for direct subtraction (but not for addition) (Thompson, 1999). However, Benz (2005) documented only small differences in solution rates to additive operations in grade 2 in this context. Moreover, the same effect showed for the number of correctly identified operations. In this study, students selected an additive model in 18.4% of the responses, when subtraction was the correct model. In 11.6% of the responses to tasks, for which the correct mathematical model was addition, a subtractive model was chosen. This could be interpreted, albeit very weakly, as a preference of the investigated learners for additive models over subtractive models. Literature on proportional reasoning discusses preferences for additive-subtractive or multiplicative-proportional models as one explanation, why some learners solve word problems on proportional reasoning with additive-subtractive mathematical models, but other learners use proportionality strategies (Modestou & Gagatsis, 2013; Resnick & Singer, 1993; Van Dooren, De Bock, Vleugels, & Verschaffel, 2010). It seems plausible that learners choose operations according to such a preference, unless they can make a clear decision for a particular operation based on how they understand the presented situation. If such a preference occurs in favor of addition in this context, this could be due to a perceived lower difficulty of addition problems compared to subtraction problems. Introducing addition earlier than subtraction in the context of number decomposition could also explain why some learners prefer addition. It would also be worth investigating whether an unbalanced ratio of addition and subtraction tasks in textbooks or in the classroom could be a possible cause. Independently of this, it is conceivable that enhancing understanding regarding the symmetry of relations and actions could reduce such a preference effect.

In summary, these results on the first research question show that solution rates in word problems clearly depend on which features of the situation structure are used to represent a mathematical structure in the word problem. Thus, regardless of the fact that some classical results could not be replicated for the tested sample, the question arises, if enriching the situation model by alternative features of the situation structure during the reading process could have positive effects on word-problem solving.

4.5.2 Q2: Using situation structures as a spontaneous solution strategy

In the literature, approaches suggest that learners could enrich their individual situation model either by alternative semantic structures (Greeno, 1980) or by alternative views on quantitative comparisons (Stern, 1993) as a solution strategy for additive one-step word problems. Practice-oriented contributions to mathematics education also suggest corresponding strategies (Fromme et al., 2011). The second research question was examined in preparation for developing an intervention program, which aims at enhancing flexibility in dealing with arithmetic situations for supporting learners with word-problem solving. It was investigated whether learners can already recognize the structural similarities between different verbally presented semantic structures without support and use them for successfully solving word problems (goal 2.1). If this was the case, an intervention program would have to aim less at conveying the relationships between different situation structures (here especially between semantic structures), but rather at stimulating using this knowledge as a reading and problem-solving strategy. Given the unexpectedly small differences in difficulty between semantic structures in this preliminary study, the results on this research question are not surprising. There is hardly any evidence that processing a structurally similar situation beforehand is recognized by learners spontaneously or because of a short hint, and used to solve a subsequent word problem.

This could have several reasons: (1) Based on the results for the first research question, it is possible that learners do not find this strategy helpful, since the two tasks of a word problem pair are not substantially different in difficulty. This would speak for addressing flexibility with respect to other information encoded in the situation structure of word problems, such as the a/s wording of relations and actions. (2) It may be that learners are capable of using structural similarity but do not use this knowledge because of socio-mathematical norms about the solution of word problems (Yackel & Cobb, 1996). The observation that an explicit hint to use structural similarity between tasks had no effect speaks against this explanation. (3) It could be that learners do not use this strategy for other reasons, although they would be able to do so. It would then be an open question, which reasons these could be and whether a corresponding intervention can stimulate or intensify using the strategy. (4) It is possible that learners of this particular age are not able yet to develop adequate flexibility in dealing with arithmetic situations. Any results (e.g., from developmental psychology) indicating such a fundamental limitation are not known. On the contrary, past research in developmental psychology has documented very early abilities in many domains for requirements that are actually assumed to be "abstract", when they are embedded in concrete and accessible contexts (as in the word problems considered here) (e.g., Koerber, Mayer, Osterhaus, Schwippert, & Sodian, 2015 for scientific reasoning; Markovits & Thompson, 2008 for logical reasoning). In the end, however, the question of whether a strategy can be learned cannot be answered with observational studies, but only with the help of intervention studies. (5) It is also possible that the learners simply do not have the necessary flexibility in dealing with arithmetic situations to infer the structural similarity of the tasks from the text base. In this case, it would be an essential question, whether the envisioned intervention is suitable to stimulate such flexibility and to encourage using flexibility for solving word problems. In order to separate the last two explanations, it would be helpful to be able to measure flexibility in dealing arithmetic situations independently from word problems. A first instrument for this purpose is already available from a thesis (Weber, 2016) and has been successfully pilot-tested. Results of this thesis indicate that some learners are able to recognize the structural similarity of differently described additive situations. In particular, this weakens any objections to the learnability of this flexibility.

Given the small differences in the difficulty of word problems on different semantic structures and the results on models of number concept acquisition (Fritz et al., 2018), it seems plausible that many learners also are proficient in the challenging relational number concept underlying the compare problems by the middle of the second grade. The results for the first research question suggest that flexibility in dealing with symmetrical relations and actions is a particularly significant challenge at this point in development. This developmental step is not systematically described in previous models of number concept acquisition (e.g., Fritz et al., 2018). If an intervention that stimulates this flexibility has effects on word-problem solving skills, this would suggest that understanding the symmetry of relations and actions is still developing at this stage. It would still be an open question, whether such understanding can be developed jointly for all semantic structures, or whether the development starts earlier for certain (e.g., dynamic) semantic structures than for others (e.g., static structures related to the relational number concept).

4.5.3 Limitations

Due to its design, the present study can certainly not provide final answers to many questions. While the sample is suitable for identifying medium to large effects of task features, it is certainly not sufficient to clarify small differences in detail. However, in order to substantiate the design of an intervention, the mechanisms associated with larger differences in learner performance are also of primary interest. Nevertheless, it cannot be concluded from this study that some of the considered factors do not have a certain relevance for solving word problems, which, however, could not be identified in the form of significant effects with the chosen sample size.

Due to its experimental design, the study can make statements about causal relationships between task features and task difficulty, but questions remain unanswered. While this study can provide valuable information for designing an intervention, it certainly cannot say anything about its potential impact.

It should also be considered, which contexts and word problem types were selected when interpreting the results. The aim of the study was to investigate how learners cope with certain demands when dealing with verbally described additive one-step word problems. The range of situations considered was therefore deliberately limited. For example, no combine problems were included, since these are difficult to combine with compare problems. Furthermore, only one-step word problems were considered. How learners apply their skills and their knowledge of additive situations in more complex, possibly more authentic situations goes beyond the focus of this study.

4.5.4 Conclusion and implications

This study takes a specific look at the role of language in learning mathematical concepts. Flexibility in dealing with arithmetic situations when solving additive one-step word problems describes the ability of learners to enrich their individual situation models by inferences with additional features on the level of situation structures (Fig. 1). Thus, a theoretical concept is proposed that appears to be largely language-independent at first. However, the proposed flexibility may be essential for understanding linguistic information about situations that are described mathematically. The concrete text base of such situations usually highlights only a part of the situation structure's features, while other features and alternative views remain hidden and can only be added by making inferences. How the required flexibility relates to concepts of subject-specific language skills such as mathematical vocabulary (Peng & Lin, 2019; Powell, Driver, Roberts, & Fall, 2017; Schindler et al., 2019; Ufer & Bochnik, 2020) and skills for mathematical text comprehension (Bochnik & Ufer, 2016; Ufer & Bochnik, 2020), to general language skills, or to less language-related parts of conceptual knowledge would need to be clarified in further studies. Furthermore, another interesting follow-up question would be to what extent such increasing flexibility contributes to constructing an interlinked knowledge structure on additive situations, which then integrates all word problem types in the sense of an "abstract meta-schema".

In addition to the above-mentioned consequences for the planned intervention, the study yields several essential implications.

(1) The differences in difficulty between different semantic structures described in the previous, partly older, literature in first grade (e.g., Stern, 1998) cannot be fully replicated at this stage for the tested sample in the middle of second grade. Although researchers indicate that a quarter of learners have developed an understanding of relational numbers by the beginning of second grade (Fritz et al., 2018), it is noteworthy that this no longer seems to be a real obstacle for the word problems used in this study by the middle of second grade. The initial assumption that reinterpreting different semantic structures could be particularly helpful is subsequently not supported by the study for the time being.

(2) Instead, the study shifts the focus on understanding the symmetry of relations and actions. Although it seems plausible at first that this symmetry becomes particularly relevant in compare problems (Fuson et al., 1996; Stern, 1993), the results indicate that corresponding differences in difficulty are not significantly weaker for the other two semantic structures considered. This

suggests that learners develop an understanding of the symmetry of relations and actions only after acquiring the relational number concept. No evidence was found that this occurs more easily or earlier for certain semantic structures than for others. Overall, this speaks in favor of giving equal attention to all three considered semantic structures, and thus especially to dynamic structures, when dealing with linguistic means for describing quantitative comparisons. With regard to the planned intervention, the study thus suggests that interventions should address the described symmetry with regard to all three considered semantic structures, and in this sense, encourage a flexible approach to mathematical situation structures.

(3) The study suggests that learners prefer additive models over subtractive models. Various reasons for this are conceivable, ranging from students' individual assessment of operations (e.g., perceiving addition as easier than subtraction) to explanations regarding the learning opportunities. Thus, it should be examined to what extent teachers address situations on additive models with equal frequency and simultaneously in the classroom. If students systematically prefer certain operations, independent of the task, this may indicate problems in understanding additive situations. However, in case of unfavorable task selection, it is also possible that the illusion arises for learners that they comprehended the situation, if they succeed frequently with the preferred strategy.

(4) Finally, the study does not provide any evidence that students use structural similarity of directly successive tasks spontaneously. Nor does it show any evidence that this would merely be an effect of socio-mathematical norms in the classroom context (Yackel & Cobb, 1996). Although using such structural similarities is certainly not a central goal of learning mathematics, it does indicate that developing and using flexible knowledge about additive situations cannot be considered a resolved issue in mathematics education. Further research should investigate, whether and under which conditions learners can draw on similarities between additive situation structures to solve word problems.

To understand the mechanisms behind the identified and unidentified correlations on effects, a detailed analysis of individual problem-solving and learning processes is necessary, as could be done, for example, in the context of an intervention study. It would also be interesting to reconstruct cognitive processes using qualitative analyses. With the help of the "thinking-aloud" method (Lewis, 1982), the cognitive processes and approaches of learners when working on the word problem pairs could be explored.

In summary, the preliminary study examines and systematizes older findings on additive one-step word problems. It reveals a more differentiated picture than the one anchored in the literature review (see Chapter 2.4.2). With regard to the planned intervention, the study contributes essential information how the intervention could be conceptualized. The hypothesis that flexibility in dealing with arithmetic situations is already used to a significant extent could not be confirmed. The effects of fostering this flexibility can only be investigated in an intervention study. Beyond that, the study provides evidence that a – theoretically quite desirable – flexible use of situation structures is not systematically applied as a strategy to cope with demands in word problems. This suggests that possible obstacles to using this strategy should be investigated.

5 Intervention study

5.1 Introduction

Based on the results of the preliminary study, an intervention study was conceptualized. The goal of this intervention study was to examine the feasibility of fostering flexibility in dealing with arithmetic situations with the suggested strategies (Greeno, 1980; Stern, 1993) and its effect on word-problem solving. The intervention program considers previously reported results (see Chapter 4.4.1) by also putting emphasis on the symmetry of relations and actions. Nonetheless, for developing flexibility in dealing with arithmetic situations, not only the symmetrical statements, but also the semantic structures should be interlinked (see also Fig. 3 for interlinked statements on one situation). By systematically integrating different descriptions into a network of linked perspectives on arithmetic situations, learners can advance their conceptual knowledge on such arithmetic situations (Scheibling-Sève et al., 2020). Therefore, the intervention study investigates the effect of developing both the Inversion Strategy and the Dynamization Strategy on the students' flexibility (goal 2.2) and their word-problem solving skills (goal 2.3). Since the study was conceptualized as a "feasibility study", students were tutored additionally to regular math lessons. Other students, who did not receive tutoring, formed the control group. To reduce the influence of other factors on the intervention, no actual word problems were solved during the intervention (in terms of setting up an equation and calculating the result). Instead, the activities focused on the level of the situation structure and corresponding verbal descriptions: Students analyzed, compared, and produced various descriptions of arithmetic situations that emphasized different features of the situation structure. These activities put high demands on using language cognitively and communicatively (see Chapter 2.1.2, Chapter 2.5.3.2). Thus, this intervention study also investigates, to what extent language skills influence students' development of flexibility and their word-problem solving skills (goal 2.4). By collecting process data during the intervention sessions, the study also strives to gain detailed insights into the students' development of flexibility and the characteristics of this new construct (goal 2.5).

In the following chapter, the design of the intervention study will be presented. Within this chapter, the procedure of the entire study will be outlined. It will be reported, which test instruments were used in the pretest and posttests and how the participants were assigned to experimental group and control group. In the end, the intervention program and the process data collected during the intervention will be described. Subsequent to this chapter, first a quantitative, then a qualitative analysis will be presented in Chapters 5.3 and 5.4.

5.2 Design

5.2.1 Overview

The intervention study was conducted with *N* = 130 second graders from ten classrooms in three primary schools located in Munich, Germany. The students first answered a pretest in the second half of the school year (March/April 2019), which was directly followed by a five-week intervention with only a part of the sample (May 2019, for selection procedure see Chapter 5.2.3). Directly after the intervention, all students participated in a posttest (June 2019) and a follow-up test four weeks later (July 2019) to investigate the long-term effect of the intervention on the students' skills. Fig. 9 gives an overview of the intervention study.

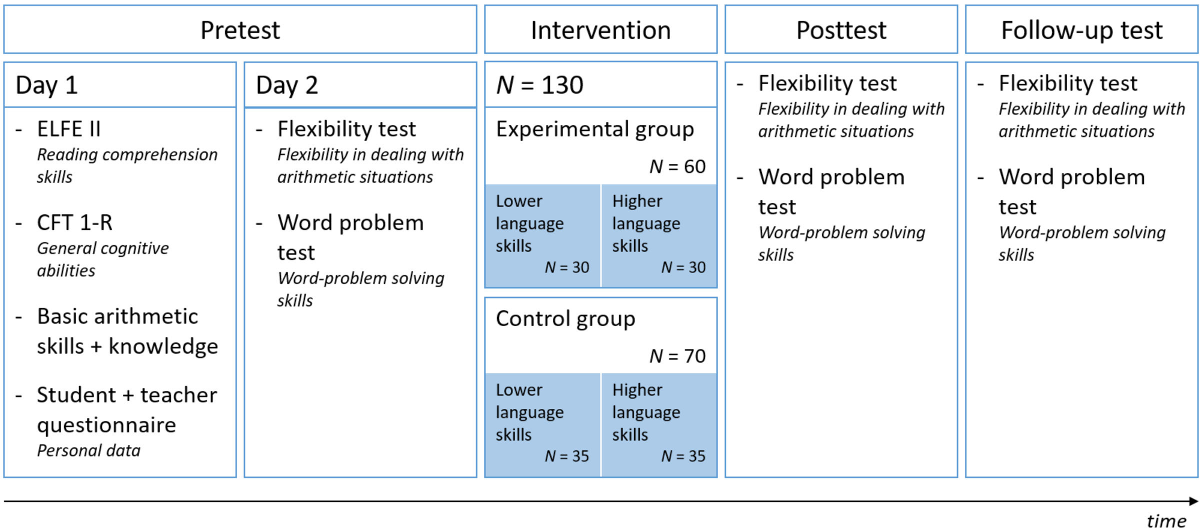

Fig. 9: Overview of the intervention study

5.2.2 Test instruments

The test instruments were designed as paper-pencil tests, which the students completed as individual work. Trained university student assistants administered the tests in each classroom. For comparability, the administrators followed guidelines with pre-formulated instructions. Preliminary to each single test instrument, they solved sample tasks with the students and answered emerging questions. After phases of 15-25 minutes, the learners took short relaxation breaks and participated in physical activities.

The whole sample answered the tests on flexibility in dealing with arithmetic situations (in the following, “flexibility test”) and word-problem solving skills (in the following, “word problem test”) during all three measurements. This intended to investigate the sample's performance growth in these two scales. In the pretest, also personal data, students' basic arithmetic skills and knowledge, their reading comprehension skills (as an indicator for language skills), and their general cognitive abilities were collected as control variables. Students' answers were coded based on a coding manual. In the following sections, the applied test instruments will be introduced.

5.2.2.1 Flexibility in dealing with arithmetic situations

To measure the construct of flexibility in dealing with arithmetic situations, a test instrument was newly developed within a thesis (Weber, 2016). This instrument was already pilot-tested successfully in the preliminary study and adapted based on gained experience. To measure the pursued flexibility, the items were embedded into a story about two twins, who tell the learners about a birthday party they visited (see Fig. 10).

Fig. 10: Sample item[11] to measure flexibility in dealing with arithmetic situations

The learners were asked to decide, if the statements of the twins are equivalent or not. The twins' statements emphasized different features of the situation structure (see Chapter 2.2.2). However, some statements did not refer to the same situation (as in Fig. 10). Statements that were indeed equivalent corresponded either with the Dynamization Strategy by contrasting different semantic structures, or the Inversion Strategy by contrasting different a/s wordings. For the Dynamization Strategy, statements on the comparison of sets were either contrasted with change statements or with equalize statements. Statements on combining two sets were omitted, since no authentic combination with situations on the comparison of sets could be created. For the Inversion Strategy, statements on the same semantic structure, but with varying a/s wording were contrasted. Again, statements on combining two sets were omitted, since the a/s wording cannot be varied for this semantic structure.

As the results from the preliminary study indicated that a stronger focus on the a/s wording was necessary (see Chapter 4.4.1), some of the original items on the Dynamization Strategy were replaced with items on the Inversion Strategy for the intervention study, which resulted in eight items on dynamization and twelve items on inversion.

The students had 15 minutes to answer 20 items. With the additional option to check the box "I don't know", it was intended to avoid that students guessed the answer in case they were uncertain. If students finished before the time ended, the administrators encouraged them to reconsider the items, for which they picked "I don't know".[12]

[11] Items were translated from German for this dissertation.
[12] In the pretest, students picked "I don't know" in 4.5% of all cases (posttest: 3.0%, follow-up test: 2.5%).

Students answered the test during all three measurements. The answers were scored dichotomously, with missing answers and the option "I don't know" treated as incorrect answers. For reliability analysis, Cronbach's alpha was calculated for each measurement. The test instrument's reliability was satisfying (pretest: $\alpha = .80$, posttest: $\alpha = .82$, follow-up test: $\alpha = .88$). In the pretest, the participants scored $M = 14.04$ points on average out of 20 total points, with a standard deviation of $SD = 4.12$ (posttest: $M = 15.52$, $SD = 4.14$, follow-up test: $M = 16.06$, $SD = 4.32$). In particular, the values of the follow-up test indicate ceiling effects.

5.2.2.2 Word-problem solving skills

The test instrument for measuring word-problem solving skills was newly created for the preliminary study (see Chapter 4) and adapted afterward. The items were based on word problems used by Stern (1998). Researchers usually investigated the difficulty of 14 to 16 different types of word problems: six on comparison, six on change, two on combination, and occasionally, two on equalization (e.g., Kintsch & Greeno, 1985; Nesher et al., 1982; Riley et al., 1983; Stern, 1994). By systematically varying the semantic structure, a/s wording, and unknown set, the collection of 16 word problems by Stern (1994) was extended within this project, resulting in 20 word problem types in total (for an overview of the full collection of word problem types see Radatz et al., 1996, p. 79 f.). Stern (1994) already investigated two equalize problem types with unknown change set. Four equalize problems were added to the test instrument, in which the change set is *given* (see Tab. 1). This extension makes it possible to match the different types of compare problems with their dynamic counterparts. As far as is known, no study has investigated the difficulty of these four equalize problems yet.

Features	Compare problems	Equalize problems
Unknown difference / change set Add. wording	Susi has 4 marbles. Max has 7 marbles. How many marbles does Max have more than Susi?	Susi has 4 marbles. Max has 7 marbles. How many marbles does Susi have to get to have as many marbles as Max?
Unknown difference / change set Sub. wording	Susi has 4 marbles. Max has 7 marbles. How many marbles does Susi have less than Max?	Susi has 4 marbles. Max has 7 marbles. How many marbles does Max have to put away to have as many marbles as Susi?
Unknown compare / result set Add. wording	Susi has 4 marbles. Max has 3 marbles more than her. How many marbles does Max have?	*Susi has 4 marbles.* *If she gets 3 marbles, she has as many marbles as Max.* *How many marbles does Max have?*

Unknown compare / result set Sub. wording	Max has 7 marbles. Susi has 3 marbles less than him. How many marbles does Susi have?	*Max has 7 marbles.* *If Max puts 3 marbles away, he has as many marbles as Susi.* *How many marbles does Susi have?*
Unknown reference / start set Add. wording	Max has 7 marbles. He has 3 marbles more than Susi. How many marbles does Susi have?	*Max has 7 marbles.* *If Susi gets 3 marbles, she has as many marbles as Max.* *How many marbles does Susi have?*
Unknown reference / start set Sub. wording	Susi has 4 marbles. She has 3 marbles less than Max. How many marbles does Max have?	*Susi has 4 marbles.* *If Max puts 3 marbles away, he has as many marbles as Susi.* *How many marbles does Max have?*

Tab. 1: Sample compare problems based on Stern (1998) and their dynamic counterparts. The equalize problems in cursive characters were added through systematic variation

The sample word problems in Tab. 1 contain the same subjects, involved objects, and number material to demonstrate the respective counterparts. However, for the test instrument, these features were varied for each word problem type. Each of the twenty word problem types was realized in one item (see Fig. 11 for a sample item).

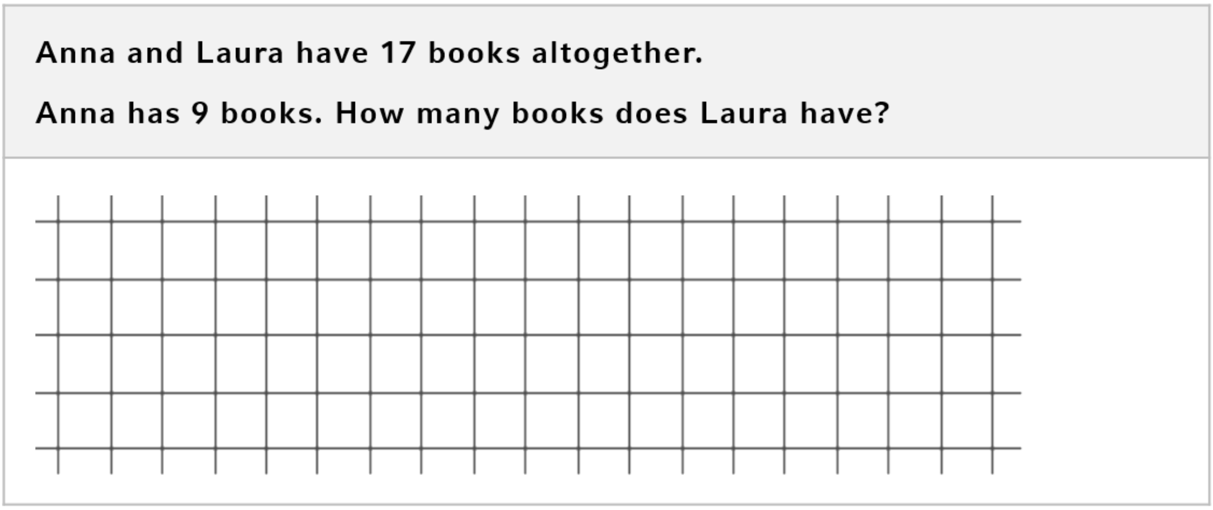

Fig. 11: Sample item to measure word-problem solving skills (combine problem with unknown subset)

The depicted situations either dealt with two persons possessing certain objects (e.g., Susi has a certain number of books) or with the composition of two subgroups (e.g., tulips and roses) within the same category (e.g., flowers). To minimize the linguistic complexity as far as possible, the word problems contained only relevant information, basic vocabulary, and similar sentence structures. The involved numbers were chosen in a way, which did not make it possible to recognize a pattern, such as “big number and small number in a text requires subtraction, two small numbers require addition” (as described for division in Páchová &

Vondrová, 2021). Only numbers up to 30 were used to limit the arithmetic demand, and only carry/regrouping operations (arithmetic operations including the change of a tens digit, see Chapter 2.4.3) were chosen to ensure comparability. Equalize problems were formulated in a way that eliminated ambiguity as good as possible, so that the idea of equalizing *one* set instead of both sets became clear. This was implemented by using verbs as in the following example, which did not allow interpreting the situation as an equalization of both sets: "Susi has 13 cookies. If she *eats* 5 cookies, she has as many cookies as Max. How many cookies does Max have?"

The test was implemented in a multi-matrix-design: Each student only worked on ten items of the entire collection of word problems. This aimed at reducing the second graders' workload. To ensure that the booklets were similarly difficult, the twenty word problems were divided into four different subparts (A, B, C, D) of comparable difficulty with five word problems each. Each booklet contained two of the four subparts. The systematic combination of all subparts resulted in six different booklets in total (AB, AC, AD, BC, BD, and CD). Administrators instructed the students to write down the entire calculation (e.g., 17 – 9 = 8), but to omit an additional answer sentence as students might be used to from regular mathematics lessons. Students had twelve minutes for solving the word problems. The coding procedure was the same as in the preliminary study (see Chapter 4.3.3). The scores were scaled with a one-dimensional Rasch model. The WLE reliability across all three measurements was .63. The mean item difficulty across all three measurements was -1.21 with a standard error of 0.20, which indicates a relatively low difficulty of the test instrument in general.

5.2.2.3 Basic arithmetic skills and knowledge

To measure the students' basic arithmetic skills and knowledge, a test instrument from the LaMa project with third graders (Bochnik, 2017) was adapted for second graders. The test includes context-free tasks with short, simple instructions. At first, the students answered basic calculation tasks, which rather relate to technical skills in the context of fluently adding and subtracting numbers ranging until 100 (with and without carry/regrouping operations). Then, the tasks progressed to more demanding formats requiring conceptual basic arithmetic knowledge, for example on the relationship between mathematical operations (e.g., by asking for all four calculations that can be conducted with the numbers 7, 8, and 15), the meaning of the equal sign (e.g., by completing an equation such as 5 + _ = 11 – 2), or on place value (e.g., by identifying numbers that consist of the same amount of tens and ones). The students had 15 minutes to complete the 16 tasks.

Students completed the tasks at Day 1 of the pretest. The students' answers were scored dichotomously, with missing answers treated as incorrect answers. For reliability analysis, Cronbach's alpha was calculated. The test instrument's reliability is satisfying (α = .82). On average, the students scored M = 7.49 points out of 16 total points with a standard deviation of SD = 3.80.

5.2.2.4 Language skills

One research question of the dissertation project (see goal 2.4 in Chapter 3.2.2) was how language skills influence the intervention's effect. Following empirical studies (e.g., PISA, Baumert et al., 2001) that investigated the relation of (general) language skills to mathematics performance and development by using tests of reading comprehension (Heinze et al., 2007; Peng et al., 2020; Ufer & Bochnik, 2020), the students' language skills were also measured with a reading comprehension test, namely the German test ELFE II (Lenhard & Schneider, 2018). This test provides the opportunity to assess receptive language skills based on reading fluency and accuracy with a larger sample. Norms exist for each grade and the exact month of the school year. This facilitates comparing the students' scores to the norm sample of second graders during the seventh month of the school year. To follow the given recommendation in the test manual, the test was restricted to items that were considered appropriate for second graders. Thus, students only answered items on the word level (choosing one out of four words which matches the given picture) and sentence level (choosing one out of four words that fits into the given sentence), but not on the text level (choosing one out of four possible answers to a question on a short text). As instructed in the test manual, students worked three minutes on each subtest. Their answers were coded dichotomously and in accordance with the instructions of the test manual. In the end, the combined score in both subtests was used to determine the percentile rank with respect to the norm sample. On average, the students achieved M = 45.03 raw points with a standard deviation of SD = 15.5, which is in line with the average performance of the norm sample of the test. For reliability analysis, Cronbach's alpha was calculated. The test instrument's reliability is excellent (α = .97). The test manual reports an excellent odd-even-split-half reliability with r_{tt} = .96.

5.2.2.5 General cognitive abilities

The students' general cognitive abilities were measured with the subscales "Similarities", "Classifications", and "Matrices" of the Culture Fair Intelligence Test CFT 1-R (Weiß & Osterland, 2013), a language-free intelligence test for children aged from five to eleven years. Based on the theory of Cattell (1987), these subscales measure characteristics of fluid intelligence, such as problem solving and reasoning, in a culturally fair setting. This way, the students' scores could be used to control for general cognitive abilities with a minimized influence of language and culture. The implementation followed the instructions in the test manual.

Students completed the tasks at Day 1 of the pretest. The students' answers were scored dichotomously, with missing answers treated as incorrect answers. For reliability analysis, Cronbach's alpha was calculated. For all three subscales, the reliability is acceptable (subscale "Similarities": α = .66; "Classifications": α = .73; "Matrices": α = .80). The reliability of the three subscales combined is poor (α = .56), however, this may be because only three of six subscales were used. For the entire test instrument with all six subscales, the test manual reports an excellent Kuder-Richardson (Formula 8) reliability (${}_{bat}r_{(tt)}$ = .97). On average, the students scored *M* = 30.41 points out of 45 total points, with a standard deviation of *SD* = 5.98.

5.2.2.6 Personal data

To collect personal data of the participants, students and their teachers were consulted at Day 1 of the pretest. The students answered one item on their family's socio-economic status. As an indicator for socio-economic status, the "books-at-home-index" (Paulus, 2009) was used. The widely used item provides an opportunity to collect this kind of data from young children. Students were asked to estimate the amount of books their families own. There were five answering options, ranging from "No or almost no books (0-10 books)", over "One bookshelf (11-25 books)", "One entire bookcase (26-100 books)", "Two entire bookcases (101-200 books)", to "Three entire bookcases or more (over 200 books)". Pictures of the bookshelves and bookcases supported the learners with estimating the quantities correctly.

The students' teachers answered the other items on personal data to save time and reduce the number of questions for the students. Another reason to consult teachers was that they could respond to questions on matters that might be difficult to answer for second graders. To gather information on the students' language and migration background, the teachers named the students' family language(s) and the period of time they have lived in Germany yet. As further background information regarding the students' education, their teachers were asked about the learners' participation in additional educational support, such as reading support. Additionally, teachers provided data on the students' gender and their age at Day 1 of the pretest.

5.2.3 Group assignment

After the pretest, the students were randomly assigned either to the experimental group, who received the intervention, or to the control group (see Fig. 12). To control for language skills, the students' scores in the ELFE II reading comprehension test were considered. First, the students' scores were ranked within each class. Then, three pairs of students were formed starting from each end (highest scores and lowest scores). By tossing a coin, one student of each pair was assigned to the experimental group, and the other student to the control group. This procedure intended to generate intervention groups with heterogeneous language skills, but also avoid that the control group only consisted of students with average language skills. Heterogeneous groups were chosen to preserve authenticity (as in regular classrooms). Furthermore, learners with lower language skills may benefit from learners with higher language skills (Pyle, Pyle, Lignugaris/Kraft, Duran, & Akers, 2017), as they could not only use the tutor's language as a resource for developing flexibility, but also benefit from examples that were set by peers (Schneeberger, 2009). Forming heterogeneous groups of students with lower or higher language skills may also facilitate to investigate how language skills influence the intervention's effect (goal 2.4). The remaining students, who were not involved in the six pairs of each class, were also assigned to the control group. This way, a control group (*N* = 70) and an experimental group (*N* = 60) consisting of ten intervention groups of six students each (one per classroom) were formed.

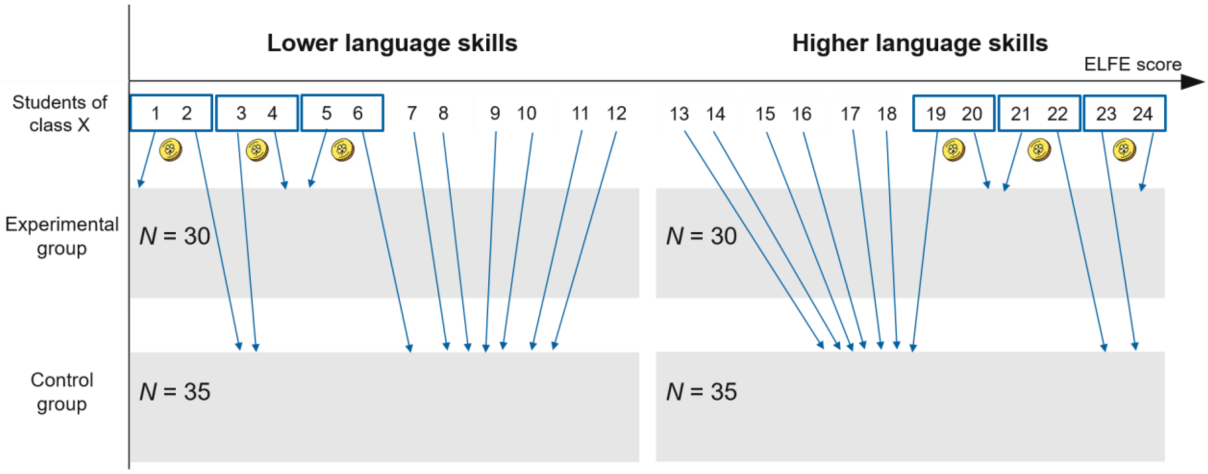

Fig. 12: Overview of the procedure how students were randomly assigned to experimental group and control group based on their score in the ELFE II test (showing a sample procedure for class X)

5.2.4 Intervention program

5.2.4.1 Overview of the program

Organization. The intervention comprised ten 40-50 min small-group sessions over five weeks. Both the experimental group and the control group received regular mathematics lessons during the five weeks, and the experimental group received the intervention *additionally* to regular mathematics lessons. This setting was chosen to examine, if the pursued flexibility in dealing with arithmetic situations can be fostered at all. To ensure that both groups still participated in a comparable amount of regular mathematics lessons, students of the experimental group were mostly tutored during other subjects than mathematics in a different room.

Pre-service teachers acted as tutors and were instructed beforehand in a preparatory meeting. As a guideline, the tutors followed an intervention script that contained information on content, procedure, and duration of the activities. The script also provided wording suggestions and instructions how students could be supported. To establish similar conditions in the different intervention groups, tutors followed a specified sequence with predetermined options to adapt to the individual needs of students. When new task types were introduced, students discussed and solved these collaboratively. Tutors shifted the responsibility to students continually, with phases of individual work and teamwork followed by joint discussions for most tasks. In a pilot intervention with *N* = 4 students from another school, the suitability of the tasks was examined beforehand.

Concept. The intervention program aimed at developing flexibility in dealing with arithmetic situations. To help learners with developing skills to enrich their situation models with features of the situation structure, the program contained learning activities regarding the Dynamization Strategy and the Inversion Strategy.

The learning activities were arranged in five phases. Fig. 13 shows how the five phases of the program were sequenced over the ten sessions. After an initial phase of familiarizing with certain *Basics* (e.g., on relational statements), the Dynamization Strategy and the Inversion Strategy were introduced implicitly during *Verifying*, *Matching*, and *Describing situations*, which constituted the three main phases of the learning trajectory (see below for further elaboration). Since research and own results of the preliminary study (Chapter 4.4.1) emphasize the difficulty of understanding symmetrical relational statements, inversion was also introduced explicitly (*Symmetry of relations*).

During the five phases, learners were encouraged to use language cognitively to organize their conceptual knowledge on the features of situation structures, for example by contrasting and analyzing different features together with the group (see Fig. 6 for an example how relational statements with different a/s wording can be contrasted). This intended to provide learners with a cognitive tool to enrich their situation models with further features of the situation structure, when solving word problems in later situations (see Chapter 2.5.3.2). Using language cognitively was intensified by using language for communication: Engaging in activities such as explaining, arguing, and justifying statements or descriptions was intended to enhance rich discourse practices (design principle P1, Erath et al., 2021). For example, learners were asked to explain differences and commonalities between different statements (e.g., two symmetrical relational statements), or to justify that a certain statement matches the given situation.

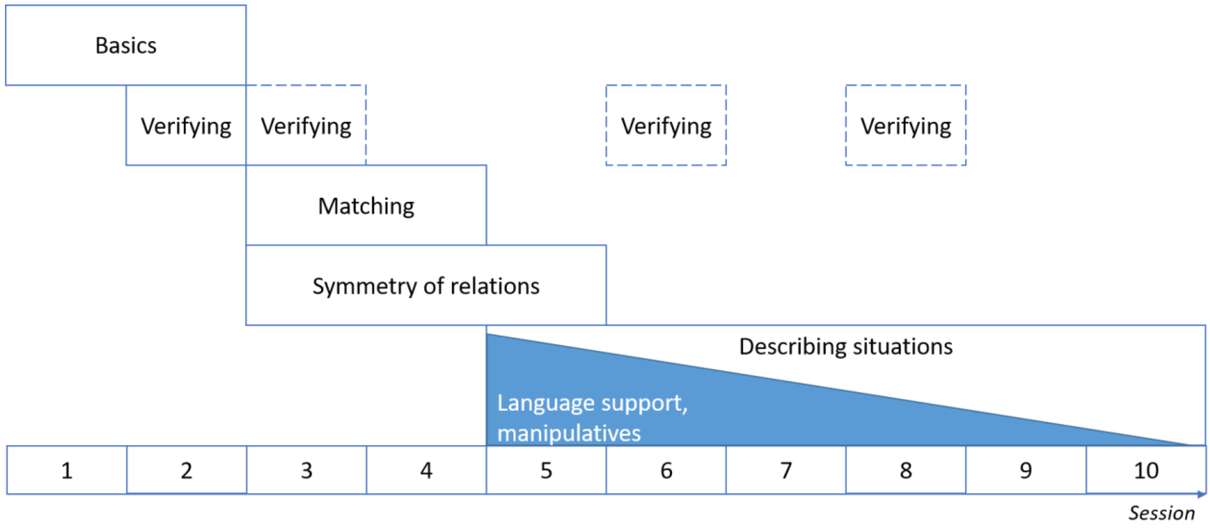

Fig. 13: Procedure of phases during the ten intervention sessions

The intervention incorporated the two instructional approaches suggested in Chapter 2.5.3. The program was based on a hypothetical learning trajectory, which progressed in difficulty in different dimensions: (1) the level of flexibility, and (2) the level of situational difficulty. (1) One way, how the learning trajectory progressed in difficulty refers to the development of flexibility. It was assumed that flexibility may progress from rather receptive flexibility (dealing with different *given* descriptions) to productive flexibility (actually describing a situation in various ways). Therefore, the students first approached the two strategies by *verifying* given descriptions of situations and *matching* given descriptions to situations (receptive flexibility), which provided an opportunity to encounter relevant linguistic means. Then, the learners were supported to transfer their knowledge on the two strategies and the linguistic means to actively *describing* situations (productive flexibility) in different ways (Fig. 13).

(2) The learning trajectory also progressed, as far as the difficulty of the provided situations is concerned. The situational difficulty depends on the features of the situation structure, which were realized in a word problem (see Chapter 2.2.2). Since the intervention focuses on interconnecting descriptions of arithmetic situations by applying the Dynamization Strategy and the Inversion Strategy, the program emphasized situations on the comparison and the equalization of sets in particular. Other semantic structures, such as changing a set or combining two sets, were not integrated intentionally, but they were welcomed, if students mentioned such descriptions. Guided by the difficulty levels of compare problems, which were elaborated and proposed by Riley and Greeno (1988), the learning trajectory progressed in the context of compare problems with different unknown sets (and their dynamic counterparts, see Tab. 1): Compare problems with unknown difference set were discussed first, since they contain no relational statement and two concrete sets. These were followed by compare problems with unknown compare set (consistent problems) and finally, compare problems with unknown reference set (inconsistent problems), which have shown to be particularly difficult in

prior studies and also in the preliminary study (see Chapter 4.4.1). The different compare problem types were connected with statements on the two involved concrete sets, their comparison and the equalization of one of these sets.

The situational difficulty did not only progress in the context of unknown sets, but also as far as the representation of arithmetic situations is concerned. At first, students worked with pictures (photographed situations as in Fig. 3 or Fig. 15) instead of text to exclude demands on reading comprehension skills and to focus solely on situational understanding. Since such pictures only allow to deal with problems with unknown difference set, and since students are also confronted with text when working on word problems, pictures were then replaced by written descriptions of the situation in the style of word problems. The question on the numeric solution of a word problem was omitted to avoid an effect by simply practicing word-problem solving. Besides this overall progression, connections between different types of representations (e.g., photos, texts, drawings, manipulatives, oral descriptions) were emphasized, so that learners could draw on aspects they were already familiar with and deepen their conceptual knowledge (design principle P3, Erath et al., 2021).

The intervention was designed to enhance language for learning mathematics by drawing on the design principles by Erath et al. (2021). In the following subchapters, the five phases and how the design principles were integrated in specific activities will be described in more detail.

5.2.4.2 Information on the phases of the intervention

Basics. To consolidate prior knowledge, the first phase focused on understanding (1) difference sets and (2) equalizing actions. (1) Prior research (e.g., Fritz et al., 2018; Stern, 1998) highlighted the importance of understanding that two numerical quantities (e.g., 4 and 7) differ by a third quantity (e.g., 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 to include the relation’s direction into their situation model (Fuson et al., 1996; Stern, 1998). According to Lewis (1989), determining the qualitative relationship between the two involved sets and deriving the larger quantity from this information can help students when solving inconsistent word problems. If learners do not represent the relation between two sets quantitatively, however, 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”. This has been observed in few studies (e.g., Hasemann & Stern, 2002; Mekhmandarov et al., 1996). Thus, tasks were particularly designed to link questions on qualitative and quantitative comparison, if suitable. The intervention script informed the tutors about the opportunity to link these questions sequentially by aiming at qualitative comparisons first and then discussing its quantification.

To introduce the characteristics of difference sets, tutors showed a picture of two towers with a covered base (see Fig. 14, left side), asked, which tower was higher than the other (qualitative comparison), and how many blocks the one tower was higher (quantitative comparison). The students tried to find arguments, why they could answer these questions without seeing the complete towers. Explaining the concept of difference sets as sets, which are independent from the two concrete sets in this case, may create a stimulating communication setting for building conceptual knowledge (design principle P1, Erath et al., 2021). Additionally, learners were asked to build various towers with the same difference as in the picture. This task emphasized the independence of the difference set from the two concrete sets again.

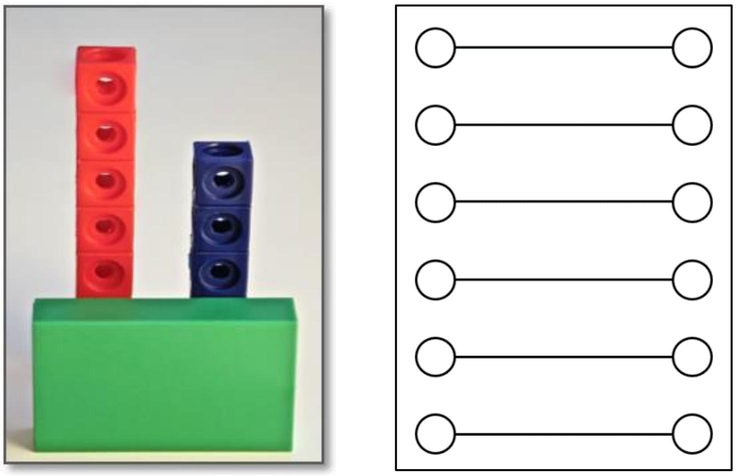

Fig. 14: Left side: towers with a covered base. Right side: template for the game *Hamstern*

(2) The Dynamization Strategy is based on equalize situations (Chapter 2.5.2). Since these situations rarely occur in textbooks (Von Damnitz, 2020), they required clarification. Before the students worked with equalize situations, the tutor discussed with them how equalizing relates to changing *one* set instead of *both* sets to ensure that learners interpret equalization as intended. Then, students played the game *Hamstern* (Verboom, 2010), which provided a context to discuss compare and equalize statements within the same situation and thus prepare the application of the Dynamization Strategy. Two students rolled a dice simultaneously and put their achieved number as chips into a template (see Fig. 14, right side). By matching the single units of the two sets one by one through lines, the template facilitated recognizing the difference set. The player with the larger amount of chips was allowed to collect the difference. The first player, who collected at least 12 chips won the round. To emphasize the connection between comparison and equalization, the tutor asked questions such as “Who has more/less chips?”, “How many chips do you have more/less?”, and “How many chips are you allowed to collect?” These questions were used to establish a language routine (design principle P2, Erath et al., 2021) in a predetermined order (qualitative comparison –

quantification – dynamization), which also appeared at other parts of the program in a similar manner.

Verifying and Matching. Before describing situations actively, *given* statements on arithmetic situations were discussed and contrasted. This provided learners with linguistic means for compare and equalize situations, which they could use for describing these situations flexibly later. Both phases contained variations of statements linked to the two strategies (see Fig. 15 for sample tasks). Analyzing and comparing these statements encouraged using 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. 15). At first, the students inspected the situation, which was either presented in a picture or in writing. For each situation, several statements on concrete sets, comparison, or equalization were provided. The students decided, if the statements were true and justified their decisions. This enhanced rich discourse practices and thus intensified students' experiences with the linguistic means (design principle P1, Erath et al., 2021). The described activities ("verify, then refine or revise", see also Chapter 2.5.3.2) were implemented as a language routine in a predetermined order (design principle P2, Erath et al., 2021), which was repeated frequently throughout the intervention. To adapt to the intervention's progress in difficulty, situations with unknown compare set and reference set were included in Sessions 6 and 8. In Sessions 3, 6, and 8, Verifying tasks were also used to track students' individual progress systematically (see also Chapter 5.2.5).

(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 in the other picture (Fig. 15). By contrasting statements on these inverse situations, this phase 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 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 the statements and situations differ (design principle P1, Erath et al., 2021) to intensify students' experiences with linguistic means for compare and equalize situations. The tutors were instructed to give impulses such as the following: "Why does this sentence fit to this picture, but not the other?" "What is the difference between the two pictures?" "There are pairs of sentences that look quite similar, but mean the exact opposite of each other. Can you find them?“

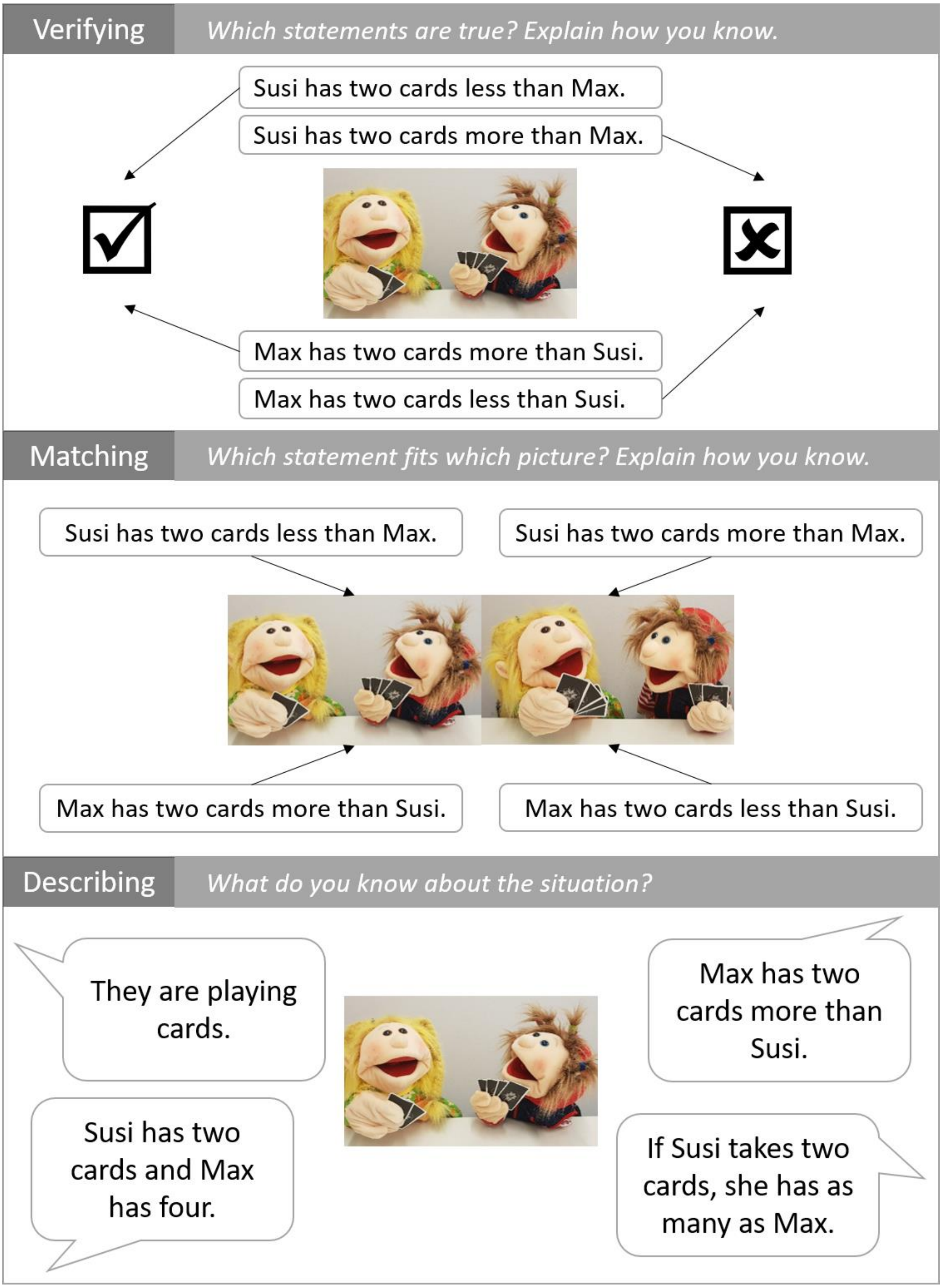

Fig. 15: Sample tasks for the three main phases of the intervention

Symmetry of relations. The purpose of this phase was to put a stronger emphasis on the symmetry of relations, since researchers assume particularly high demands of understanding this symmetry (Schumacher & Fuchs, 2012; Stern, 1993). To this end, tutors explicitly introduced the Inversion Strategy. At first, tutors provided word cards referring to relational terms and asked the learners to build random sentences by using these word cards. This macro-scaffolding was intended to help learners with activating existing resources and encourage them to formulate relational statements (design principle P5, Erath et al., 2021).

Afterward, the program approached the symmetry of relations more systematically. Tutors asked the students to invert first qualitative, then quantitative relational statements, starting with the sentence “Laura has more cookies than Klara”. At first, the tutors showed two non-transparent bags, of which one was noticeably fuller. Students guessed, which bag belonged to Laura or Klara. Then, tutors asked the students, if they could also find a different way of describing the given situation. If students struggled to find a different way of expressing this sentence, tutors chose support measures from the intervention script in a predetermined order from light to strong support (see Fig. 16). The first two measures included word cards that may encourage this inversion either with the focus on the subjects (e.g., Klara) or on the a/s wording (e.g., less). These word cards explicitly focused on providing scaffolds, which may help learners to use linguistic means needed for inversion (design principle P5, Erath et al., 2021). If neither of the word cards “Klara” or “less” helped, the tutor presented the solution and encouraged the students to compare both symmetrical relational statements (Measure 3). Students were already familiar with this activity from the Matching phase. Such an activity was intended to raise the students’ language awareness, for example, by discovering that the subjects are exchanged in the two relational statements (design principle P6, Erath et al., 2021).

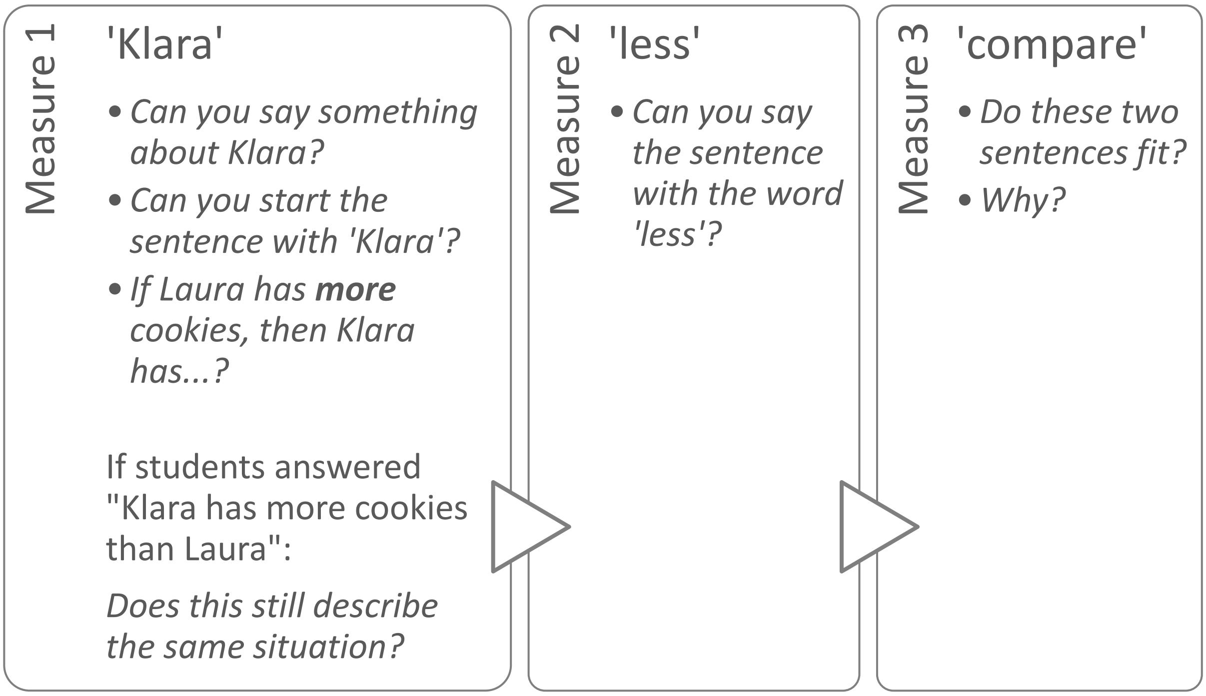

Fig. 16: Measures in a predetermined order to support learners with the inversion of relational statements (e.g., “Laura has more cookies than Klara”)

When showing the word card “Klara” during Measure 1, it was expected that some students might instead formulate the sentence “Klara has more cookies than Laura”. This sentence already contains inverted subjects, but the a/s wording remains unchanged. If this case occurred, the tutor asked, if this sentence still described the same situation and gave the following impulse: “Can you replace ‘more’ with another word, so that the sentence fits the situation?” Furthermore, tutors could resort to the bags that were provided in the beginning and remind the learners of the original situation.

To further practice the inversion of relational statements, students were asked, for example, to create statements referring to a given situation on concrete sets. These statements should include predetermined relational terms (see Fig. 17), which consisted of pairs of antonymous relational terms (less, more) to encourage students to create symmetrical relational statements. Such activities aimed to build well-connected vocabulary on relational terms and to raise language awareness by connecting statements on two concrete sets with various statements on the relation between these two sets (design principle P6, Erath et al., 2021). Besides this activity, relational terms were addressed explicitly, for example by matching pairs of antonyms, which were required for describing the a/s wording and applying the Inversion Strategy.

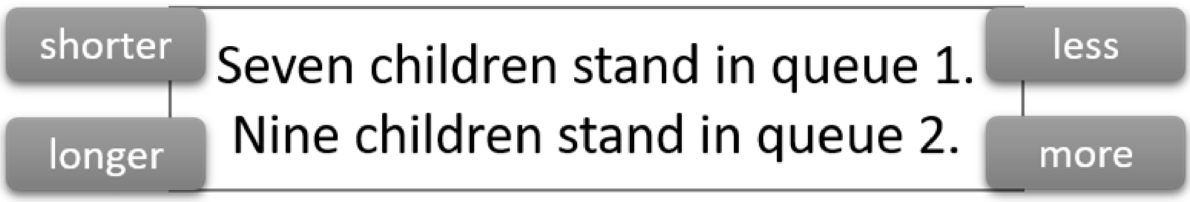

Fig. 17: Example for practicing the inversion of relational statements

The symmetry of actions was only practiced implicitly. For example, students built two towers with different numbers of bricks and described what they can do to equalize their height: either adding or taking away a certain number of bricks. The students also encountered possible descriptions of symmetrical actions, when contrasting statements during Matching or when completing sentence frames, in which an action verb was missing.

Describing situations. In later sessions, learners were also asked to actively articulate descriptions of given situations (Fig. 15). By formulating varying descriptions, learners used language cognitively to enrich their situation models. The students engaged with descriptions of situations in multiple modes (design principle P1, Erath et al., 2021): On the one hand, they generated output by talking and writing about a given situation. On the other hand, they received input by reading and listening to different descriptions by their peers and their tutor. Rich discourse practices were enhanced by encouraging flexible descriptions of arithmetic situations, as well as explanations of the differences and commonalities between different descriptions and the presented situations (design principle P1, Erath et al., 2021).

Encouraging flexible descriptions was implemented in two different ways. The first approach entailed language support through macro-scaffolding (design principle P5, Erath et al., 2021). Tutors provided sentence frames to draw attention to crucial parts of describing equalization and comparison. Students were asked to build towers of bricks and compare their height with the help of such sentence frames. For this, they either needed to fill in the blanks with one or more words (e.g., “If I …, then my tower is as high as yours.”) or choose between two options (e.g., “My tower is higher/lower than yours.”). During other activities, tutors provided sentence starters (e.g., “When I take away one brick, then …”) as a more open version of sentence frames, and word cards (e.g., “more”) to point to a certain feature of the situation structure. Tutors removed these scaffolds gradually, until the students could describe situations with a focus on comparison and equalizing without support. Another way of encouraging flexible descriptions was implemented by posing open-ended tasks. For example, asking students “What do you know about the situation?” was intended to stimulate free and flexible language use. In the final session, students were asked to write down anything that came to their minds for four situations of different difficulty. This intended to facilitate collecting further information on their flexibility at the end of the intervention and comparing it to earlier moments of the program.

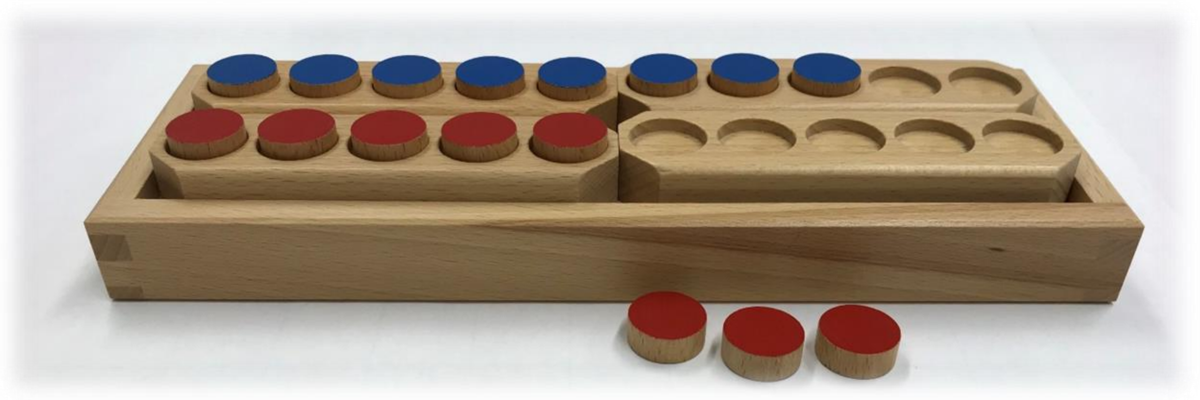

Fig. 18: *Rechenschiffchen*

Besides language support through macro-scaffolding, learners were also supported by integrating manipulatives. Representing arithmetic situations with manipulatives is strongly connected with students’ internal representations (Verschaffel, 1994). Thus, students visualized arithmetic situations with *Rechenschiffchen* (see Fig. 18), a common teaching material in German classrooms, which is similar to a twenty frame. Next to language and pictures, using this manipulative provided another opportunity to represent an arithmetic situation. Directly comparing two sets with the *Rechenschiffchen* was assumed to highlight the one-to-one correspondence and part-whole relationships, and activate conceptual knowledge on comparison (Morales et al., 1985). Students were also asked to equalize sets using the *Rechenschiffchen* to build up mental representations of equalize and compare situations and the flexible change between these semantic structures. Tutors encouraged the students to verbalize thoughts and actions when working with the *Rechenschiffchen.* This was intended to make the cognitive use of language explicit. In addition, verbalizing thoughts and actions in the context of various representations (e.g., concrete representation of two towers vs. abstract representation with the *Rechenschiffchen*) was intended to help learners with connecting different registers and representations (design principle P3, Erath et al., 2021). From Session 9 onwards, the support by the *Rechenschiffchen* was faded out to establish independence from manipulatives.

5.2.5 Overview of collected process data during the intervention

During the intervention, the tutors collected various process data. For all sessions, audio recordings were produced and transcribed afterward. These recordings intended to facilitate tracing back the circumstances during the intervention, which could help with gaining insights into students' development by analyzing transcripts of the audio recordings systematically (goal 2.5). In addition, the tutors answered a questionnaire after each session. The questionnaire contained a note field to capture information on special behavior or circumstances (e.g., which lessons the students missed, or if a student felt sick), which may be helpful to reconstruct the intervention. Furthermore, tutors evaluated each students' contributions for each session. On the one hand, it was intended to collect information to what extent students engaged with the tasks. For this, the tutor annotated for each student and for each session, to what extent they worked actively on the tasks and reacted to the tutor's questions. For single, predetermined sections, the tutor recorded with a tally list, how many contributions were made by each student to collect further information on the students' engagement during the intervention. On the other hand, the tutors assessed the quality of students' tasks. After each session, the tutor rated approximately (around 0%, 25%, 50%, 75%, or 100%), how many of the students' answers were correct. This information could be used, if further data on single students was necessary to get a richer understanding of the circumstances.

Another indicator for students' development, in particular their development of flexibility, were the Verifying worksheets. Each situation, presented as a picture or a text, formed one task, at which learners were asked to verify six different statements in the context of the situation. The situations related to different unknown sets and progressed in difficulty guided by the learning trajectory (see Chapter 5.2.4.1). The different worksheets in Sessions 3, 6, and 8 contained tasks, which were linked pairwise based on the unknown set, to track the students' development. Students worked on these tasks individually. When they finished, the tutor took a picture of the worksheet to capture the students' individual answers. Afterward, the group discussed the answers and revised incorrect statements. The six different statements per situation were associated with linguistic means, which may be helpful for describing comparison and equalization flexibly. Occasionally, number material was chosen in way, which may uncover potential misconceptions of difference sets as concrete sets. For example, if the situation was "Susi has 3 marbles, Max has 5 marbles", one challenge was to identify a statement such as "Max has 5 marbles more" as an incorrect statement. If students classified this statement as correct, this would indicate that they interpreted the difference set as a concrete set. These and all other worksheets were collected for further information on the students' answers during the program.

The intervention study served as a "feasibility study", which aimed to investigate, if flexibility in dealing with arithmetic situations can be fostered at all with the chosen approach. Therefore, students received the intervention additionally to regular math lessons. No actual word problems were solved during the intervention, 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 put up equations. This was intended to exclude effects of "simply" practicing word-problem solving. To analyze how language influenced the intervention's effect, learners were explicitly selected depending on their language skills (see Chapter 5.2.3) to form balanced groups of learners with lower and higher language skills. In the following chapter, the quantitative analyses of the intervention's effect will be outlined.

5.3 Effects of the intervention

5.3.1 Aims and research questions

The collected quantitative data were analyzed to determine the intervention's effect. For this, students' learning gain in the flexibility test and the word problem test was analyzed for the three measurements. Due to the central role of language in developing flexibility (see Chapter 2.5.3.2), it was investigated how different levels of language skills influenced the intervention's effect. Thus, the quantitative analysis focused on the following research questions:

Q1: How does an additional intervention influence students' flexibility in dealing with arithmetic situations (Q1a) and their word-problem solving skills (Q1b)? Is this effect still present after four weeks?

Based on the suggestions by Stern (1993) and Greeno (1980), the main goal of this intervention study was to foster students' flexibility in dealing with arithmetic situations with the proposed ideas (see Chapter 3.2.2, goal 2.2). Since this intervention program was implemented for the first time, the analyses aimed at investigating the *feasibility* of the chosen approach. Furthermore, it was intended to examine, if students use this ability when solving word problems (see Chapter 3.2.2, goal 2.3). A stronger increase in flexibility was expected for the experimental group compared to the control group. Moreover, it was assumed that students would be able to apply the newly developed flexibility when solving word problems. Consequently, the solution rates of the experimental group were expected to not only increase for their flexibility, but also for their word-problem solving skills.

It was also of interest, if these effects were sustainable and would persist after four weeks. Therefore, the development of the students in both scales was also investigated for the third measurement (follow-up test).

Q2: In which way do language skills influence the intervention’s effect?

Regarding the influence of language skills (Chapter 3.2.2, goal 2.4), different scenarios are plausible. Possibly, the intervention may be more effective for learners with lower language skills. Lower language skills may create a bottleneck for some learners when using language cognitively to develop and make use of flexibility in dealing with arithmetic situations (as hypothesized for second-language learners by Peng et al., 2020). The intervention was designed to support learners with lower language skills in particular to use language cognitively for the flexible description of situations. Consequently, the intervention may particularly help such learners to overcome barriers related to the cognitive use of language. In contrast, learners with higher language skills may already possess the prerequisites to use language cognitively for developing and making use of flexibility. Thus, their performance growth may not be as strong as for learners with lower language skills. It may even be possible to observe a so-called “expertise reversal effect” (Kalyuga, Ayres, Chandler, & Sweller, 2003): It is possible that instructional approaches, which are effective for novice learners, can be detrimental for experts. Processing the learning opportunities of the intervention may be redundant for experts and require more of their cognitive resources, since experts need to match their existing knowledge with the additional instruction (Kalyuga, Rikers, & Paas, 2012). These considerations speak for a stronger effect for learners with lower language skills than for learners with higher language skills.

On the other hand, the intervention may be more effective for learners with higher language skills. These learners may have better prerequisites to use language cognitively and communicatively when developing and applying flexibility, and their skills may increase more strongly. This hypothesis is in line with the Matthew effect (Merton, 1968), according to which students with a higher initial level of skill benefit more during learning further skills. However, as described in Chapter 2.5.3.2, the intervention was particularly designed to support learners with lower language skills to use the provided learning opportunities.

It is also possible that the intervention is equally effective for all students, regardless of their language skills. Prediger and Wessel (2018) reported equal effects for learners with lower and higher language skills in an intervention with seventh graders on fractions. However, further findings that investigate the requirements of learners with different language skills, regardless of their family language (language of instruction vs. other languages), are scarce (Prediger & Wessel, 2018).

5.3.2 Design and method

The original sample consisted of the students, who attended the first day of the first measurement (N = 130 students, see Fig. 9). For the quantitative analysis, only those students were included who took part at all three measurements. This allowed two trace their development from pretest to posttest and follow-up test. On the second day of the first measurement, seven of the 130 students were absent and therefore excluded. Six further students dropped out at the posttest, and four students were absent at the follow-up test. Thus, the original sample of N = 130 students was reduced to a subsample of N = 113 students (56 female, 57 male). With N = 53 students in the experimental group and N = 60 in the control group, the size of the two groups was still approximately balanced. Through the group assignment procedure (see Fig. 12), both the experimental group and the control group consisted of students with similarly distributed language skills (see Tab. 2).

	Lower language skills	Higher language skills
Experimental group	N = 26	N = 27
Control group	N = 30	N = 30

Tab. 2: The sample used for quantitative analysis, subdivided into lower vs. higher language skills (according to the ELFE II test, see Chapter 5.2.3) and experimental vs. control group

The average age of the 113 second graders was 7.7 years[13]. There were 47% of students with German as their only family language, 19% with only non-German family language(s), and 34% of students with mixed family languages (at least German and another language). While the majority of students were born in Germany (85%), 15% were born in a different country. The socio-economic status determined with the books-at-home-index (Paulus, 2009) was 3.39 points on average, with 1 being the lowest and 5 the highest value. More than every fourth student (27%) received additional support by the school, for example in reading, writing, or mathematics.

A sensitivity power analysis was conducted using the program G*Power for an F test and a repeated measures ANOVA (within-between interaction) (Faul, Erdfelder, Buchner, & Lang, 2009). Given a sample size of N = 113 and an alpha level of α = .05, an effect size of f = .15,

[13] This information is only a rough estimation of the students' age, since only their age in years was collected.

which is commonly classified as a small to medium effect (Cohen, 1992), can be identified with a power of .95.

The scores of the word problem test were scaled with a one-dimensional Rasch model. To answer the research questions related to this quantitative analysis (Q1a, Q1b, and Q2), a repeated measures ANOVA was conducted based on linear mixed models. Dependencies were modeled with a cluster sample[14] as a random factor. In addition, post-hoc tests with Bonferroni correction were conducted. The calculations were executed in R with the packages *lme4* (Bates et al., 2014) and *emmeans* (Lenth et al., 2018).

5.3.3 Results

5.3.3.1 Q1a: Flexibility in dealing with arithmetic situations / Q2

At first, the development of students' flexibility in dealing with arithmetic situations from pretest to posttest and follow-up test was investigated (goal 2.2). Fig. 19 shows the estimated marginal means in the flexibility test for students of the experimental group, who received an additional intervention (grey columns) and the control group, who only received regular lessons (black columns). The figure is split into the groups of learners with lower (left side) and higher language skills (right side) according to the reading comprehension test ELFE II (see also Chapter 5.2.3 for group assignment).

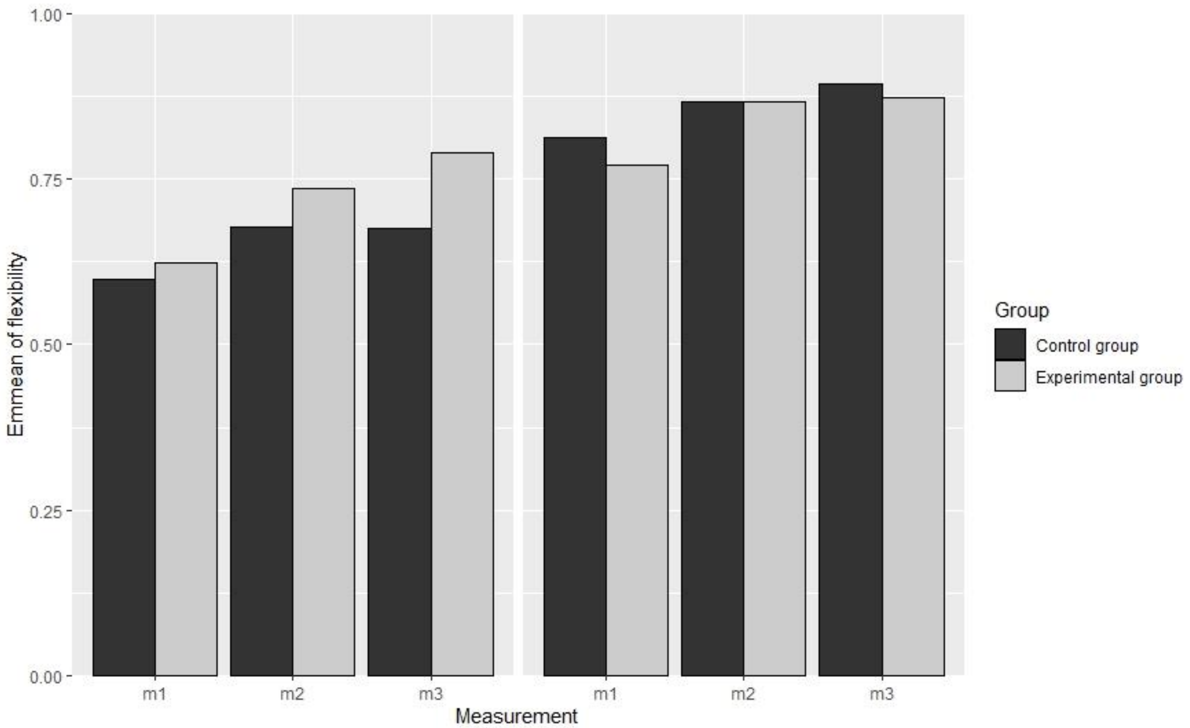

Fig. 19: Estimated marginal means ("emmean") in the flexibility test for experimental group (grey columns) and control group (black columns), split by language skills (left: lower language skills, right: higher language skills), for the three measurements (m1 = pretest, m2 = posttest, m3 = follow-up test)

[14] This refers to the students' random clustering by being affiliated to a certain classroom.

For all subgroups, there was a descriptive increase of flexibility between pretest and posttest. Only the learners of the experimental group, who had lower language skills, also seemed to increase their performance on the flexibility test between posttest and follow-up test. The performance of the other groups seemed to remain stable. A repeated measures ANOVA was conducted to test these differences for statistical significance (see Tab. 3).

	F-value	*df1*	*df2*	*p*-value	η_p^2
Language group	26.52	1	108.84	< .001***	.10
Measurement time	41.16	2	218.00	< .001***	.26
Group	0.50	1	108.97	.481	.00
Language group*Measurement time	0.82	2	218.00	.441	.01
Language group*Group	1.90	1	105.13	.171	.01
Measurement time*Group	2.66	2	218.00	.072⁺	.02
Language group*Measurement time*Group	1.47	2	218.00	.231	.01

Tab. 3: Repeated measures ANOVA for the flexibility test (Language group = lower language skills vs. higher language skills according to the ELFE II scores, Measurement time = pretest vs. posttest vs. follow-up test, Group = experimental group vs. control group).
$^{+}$: $p < .10$; *: $p < .05$; **: $p < .01$; ***: $p < .001$

Across all measurements and groups (experimental vs. control), the main effect for *Language group* revealed significant differences between students with lower and higher language skills for their performance on the flexibility test ($F(108.84, 1) = 26.52$, $p < .001$, $\eta_p^2 = .10$). Also the main effect for *Measurement time* showed significant differences in the students' performance on the flexibility test between the measurements across all groups (experimental vs. control group, lower vs. higher language skills) ($F(218.00, 2) = 41.16$, $p < .001$, $\eta_p^2 = .26$). There were no significant differences between the experimental group and the control group (*Group*) across all measurements and language groups.

The interaction effect for *Measurement time* and *Group* was not statistically significant and showed a small effect size ($F(218.00, 2) = 2.66$, $p = .072$, $\eta_p^2 = .02$). The values at least indicate an effect in tendency, according to which the experimental group and control group may have tended to develop differently along the measurements. However, this observation must be interpreted with caution. The other interaction effects, also the three-way interaction of *Language group*, *Measurement time*, and *Group*, were not significant. This indicates that – if there were marginal differences between the experimental group and the control group – the effect on students' performance in the flexibility test was not different for students with lower and higher language skills (Q2 / goal 2.4).

Post-hoc tests with Bonferroni correction allowed analyzing students' performance growth in the flexibility test in detail. The growth within the groups (experimental vs. control group) was significant for both groups and between all measurements (*experimental group*, pretest – posttest: b = 4.20, p < .001; pretest – follow-up test: b = 5.38, p < .001; *control group*, pretest – posttest: b = 2.67, p < .001; pretest – follow-up test: b = 3.13, p < .001). Then, the growth differences between the two groups were analyzed. The performance growth between pretest and follow-up test was significantly higher in the experimental group than in the control group (b = 2.25, p = .049). However, this did not show for the performance growth between pretest and posttest (b = 1.53, p = .252).

5.3.3.2 Q1b: Word-problem solving skills / Q2

It was also investigated, how students' skills in solving additive one-step word problems developed along the measurements (goal 2.3). Fig. 20 shows the estimated marginal means in the word problem test for students of the experimental group (grey columns) and the control group (black columns). The figure is split into the groups of learners with lower (left side) and higher language skills (right side).

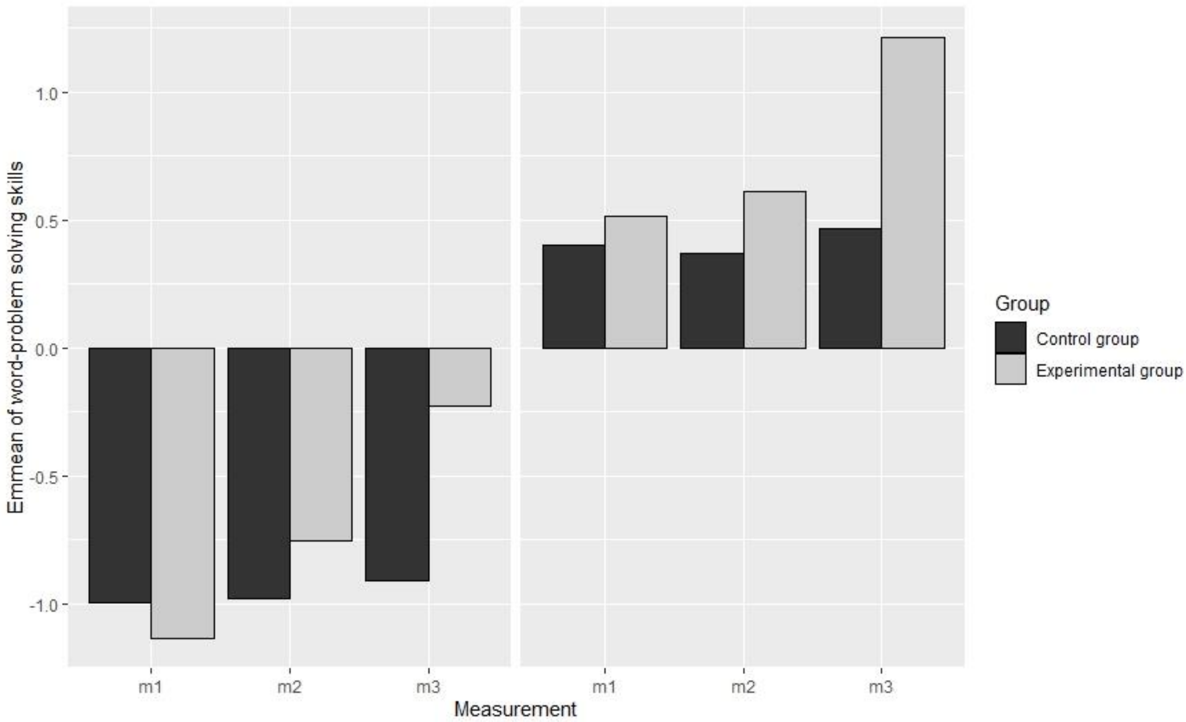

Fig. 20: Estimated marginal means ("emmean") in the word problem test for experimental group (grey columns) and control group (black columns), split by language skills (left: lower language skills, right: higher language skills), for the three measurements (m1 = pretest, m2 = posttest, m3 = follow-up test)

While the performance of the control group seemed to remain stable along the three measurements, the performance of the experimental group increased for both groups of language skills in the descriptive results. A repeated measures ANOVA was conducted to test these differences for statistical significance (see Tab. 4).

	F-value	*df1*	*df2*	*p*-value	η_p^2
Language group	35.21	1	109.00	< .001***	.13
Measurement time	4.80	2	218.00	.009**	.04
Group	1.67	1	109.00	.199	.01
Language group*Measurement time	0.16	2	218.00	.850	.00
Language group*Group	0.05	1	109.00	.816	.00
Measurement time*Group	3.21	2	218.00	.042*	.03
Language group*Measurement time*Group	0.09	2	218.00	.911	.00

Tab. 4: Repeated measures ANOVA for the word problem test (Language group = lower language skills vs. higher language skills according to the ELFE II scores, Measurement time = pretest vs. posttest vs. follow-up test, Group = experimental group vs. control group).
$^{+}$: $p < .10$; *: $p < .05$; **: $p < .01$; ***: $p < .001$

Across all measurements and groups (experimental vs. control), the main effect for *Language group* revealed significant differences between students with lower and higher language skills for their performance on the word problem test ($F(109, 1) = 35.21$, $p < .001$, $\eta_p^2 = .13$). Also the main effect for *Measurement time* showed significant differences in the students' performance on the word problem test between the measurements, across the groups (experimental vs. control group, lower vs. higher language skills) ($F(218, 2) = 4.80$, $p = .009$, $\eta_p^2 = .04$). There were no significant differences between the experimental group and the control group (*Group*) across all measurements and language groups.

The interaction effect for *Measurement time* and *Group* was significant ($F(218, 2) = 3.21$, $p = .042$, $\eta_p^2 = .03$), which means that experimental group and control group developed differently along the measurements. The other interaction effects, also the three-way interaction of *Language group*, *Measurement time*, and *Group*, were not significant. This indicates that the effect on students' performance in the word problem test was not different for students with lower and higher language skills (Q2 / goal 2.4).

Post-hoc tests with Bonferroni correction allowed to analyze the students' performance growth in the word problem test in detail. The growth within the groups (experimental vs. control group) was only significant for the experimental group between the pretest and the follow-up test ($b = 1.61$, $p < .001$), but not between the pretest and the posttest ($b = 0.48$, $p = 1$). For the control group, neither of the contrasts were significant (pretest – posttest: $b = -0.02$, $p = 1$; pretest – follow-up test: $b = 0.15$, $p = 1$). Then, the growth differences between the two groups were analyzed. The performance growth between pretest and follow-up test was significantly higher in the experimental group than in the control group ($b = 1.47$, $p = .027$). However, this did not show for the performance growth between pretest and posttest ($b = 0.50$, $p = .801$).

5.3.4 Discussion

5.3.4.1 Q1: Effect on students' flexibility (Q1a) and their word-problem solving skills (Q1b)

One aim was to investigate, how the intervention influenced students' flexibility in dealing with arithmetic situations (goal 2.2) and their word-problem solving skills (goal 2.3).

In summary, both groups made progress in the flexibility test, but the experimental group made more progress than the control group (goal 2.2). However, this effect showed only in tendency and was not statistically significant. The results on the flexibility test were unexpected, since the intervention explicitly addressed the development of flexibility in dealing with arithmetic situations. For these results, different explanations seem plausible. One possibility may be that the intervention was not suitable for fostering the students' flexibility. A qualitative analysis of students' development of flexibility during the intervention could gain more detailed insights into this hypothesis (see Chapter 5.4). Another explanation may refer to the measurement of flexibility. Possibly, the developed test instrument was only partly suitable to measure students' flexibility. It is rather unlikely that learners guessed answers, since the solution rates were located far from 50%. The relatively high solution rates and ceiling effects, especially at the follow-up test, indicate that the test was too easy (see Chapter 5.3.3.1, Chapter 5.2.2.1)). It is also conceivable that the test instrument only measured one among several aspects of flexibility. The posed tasks refer to a rather receptive aspect of flexibility (dealing with *given* statements) instead of assessing students' ability to describe situations flexibly on their own. High solution rates in the flexibility test indicate that especially learners with higher language skills may already have had an advanced level of (receptive) flexibility at the beginning of the study, which could only be increased to a limited degree.

For word-problem solving skills, the performance of the experimental group increased stronger than for the control group (goal 2.3). This descriptive performance growth becomes significant for the follow-up test. One could assume that a cause other than the intervention is responsible for the delayed growth. For example, the learners could have encountered further learning opportunities during regular mathematics lessons. However, two reasons speak against this explanation. First, the mathematics teachers were asked, which content their students worked on in between the posttest and the follow-up test. According to the teachers, all participating classrooms focused on introducing multiplication and axial symmetry in geometry. Thus, it is unlikely that regular mathematics lessons influenced the students' performance in the follow-up test. Second, due to the randomized group assignment, it is plausible that the effects can be traced back causally to the intervention. However, it remains unclear, how the performance growth after the intervention can be explained. Learners may have found the strategies they encountered during the intervention helpful. This may have triggered them to focus more on situation structures in everyday situations. Applying the strategies also after the intervention may have continuously deepened their knowledge on these strategies and the skills to make use of them. Students indeed focus on numbers and relations also outside of school, which may have given them the opportunity to process their newly gained skills even after the posttest ("SFON" and "SFOR" as terms for spontaneous focus on numbers and relations, McMullen, Hannula-Sormunen, & Lehtinen, 2013). Similar observations were made in a study by Prediger and Wessel (2018), who detected delayed, more sustainable effects for an intervention that focused on enhancing rich discourse practices.

The reported results speak against the assumption that word-problem solving skills increase due to a performance growth in the flexibility test. It appears that the intervention could not substantially increase students' performance on the flexibility test, but instead strengthened their word-problem solving skills by other means. Possibly, the intervention may rather help learners to *make use* of flexibility during word-problem solving – without explicitly solving word problems during the intervention. Accordingly, students may have already possessed a certain level of receptive flexibility, but the intervention may have helped them to develop productive flexibility and to use this productive flexibility for solving word problems.

Overall, the results indicate that fostering the application of the Dynamization Strategy and the Inversion Strategy (Greeno, 1980; Stern, 1993) by enhancing language and basing the intervention on the assumed learning trajectory (see Chapter 5.2.4.1) is a feasible way to support students. Eliciting the application of the pursued flexibility and reflecting on situation structures showed a positive effect on word-problem solving. This observation also supports findings, which emphasize how important the quality of the students' individual situation model is (Leiss et al., 2010; Stern & Lehrndorfer, 1992; Thevenot et al., 2007). It is not plausible that the effect is only caused by the additional time, since no actual word problems were solved during the intervention. However, the delayed effects for the growth in word-problem solving skills and the little effects for the growth of flexibility are still in need of explanation.

5.3.4.2 Q2: The influence of language on the intervention's effect

It was also of interest, which influence language skills have on the intervention's effect (goal 2.4). The results showed significant differences depending on language skills for both measures, but there was no indication on a differential effect of the intervention depending on language skills. Consequently, all learners seemed to benefit equally from the intervention. Prediger and Wessel (2018) have pointed out that the effectiveness of mathematics and language interventions is commonly investigated for learners with lower language skills, while findings on how learners with higher language skills benefit from mathematics interventions are scarce. The quantitative analysis reported in this dissertation contributed to gain empirical insight into this research gap by fostering learners with lower *and* higher language skills. A priori, it seemed plausible that students with lower language skills or students with higher language skills benefited stronger from the intervention (see Chapter 5.3.1). Also a similar performance growth for both groups seemed possible. The results from this study indeed support the latter assumption and match findings by Prediger and Wessel (2018), who also did not identify different demands of learners with different language skills. However, it is not clear yet, which mechanisms can explain this effect. One possible explanation may be that the first two hypotheses for Q2 are both valid to a certain extent: The intervention may help learners with lower language skills to overcome the "bottleneck" (Peng et al., 2020) and thus to increase their performance. At the same time, learners with higher language skills may be able to make better use of their resources (Merton, 1968), which may also lead to a strong performance growth. This hypothesis would have to be investigated in future research. It is difficult to prove that the intervention was equally effective for both language groups. For now, it is only known that the differences were not significant. However, the sensitivity power analysis (see Chapter 5.3.2) indicates a power of .95 for an effect size of $f = .15$, which means that the analyses should have identified larger effects. Consequently, the effect seems to be smaller. In principle, the finding that the intervention is effective for learners regardless their language skills is pleasant, since this implies that the intervention met the needs of both groups.

5.3.4.3 Limitations and outlook

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 intervention program for regular mathematics lessons. The quantitative analyses can only indicate that the intervention is effective in principle. The findings support the assumption based on suggestions by Greeno (1980) and Stern (1993) that fostering flexibility in dealing with arithmetic situations by developing the Inversion Strategy and the Dynamization Strategy is a feasible way to help learners to deal with difficult word problems.

Flexibility in dealing with arithmetic situations was measured with an innovative test instrument. Although this instrument has already been piloted in the preliminary study, the findings from the quantitative analysis of the intervention study provided new insights on the construct. Measuring flexibility with tasks that require comparing and verifying two statements seems to capture only one part of flexibility, namely receptive flexibility. The program also encouraged learners to describe situations flexibly, which seemed responsible for the performance growth in the word problem test. Based on these findings, the test instrument needs to be refined and adapted in order to capture the complete construct.

The quantitative analysis yielded information on the effectiveness of fostering flexibility with the chosen instructional approaches. However, details on *how* students developed flexibility could not be observed with this analysis. Therefore, a qualitative analysis of students' development during the intervention is necessary. This may help to understand the construct of flexibility in dealing with arithmetic situations better and to derive necessary adaptations of the program for future research.

5.4 Students' learning processes when developing flexibility

The quantitative analyses provided information on the effects of the intervention on the development of flexibility and word-problem solving skills. To gain detailed insights into the students' development, a second, qualitative analysis was conducted with four pre-selected students from the experimental group. In the following chapters, the aims of this qualitative analysis and the resulting research questions will be outlined. After introducing the sample, a short quantitative overview on process data regarding the development of flexibility will be presented. This quantitative overview underlines the necessity once again to gain detailed insights in the students' learning processes. Consequently, it was decided to conduct a qualitative content analysis (Mayring, 2014), which will be described by providing information on the coding manual and the coding procedure. Finally, the observations on the students' learning processes when developing flexibility will be reported.

5.4.1 Aims and research questions

The purpose of this qualitative analysis was to gain detailed information on the students' learning processes (goal 2.5). It intended to investigate, which learning opportunities supported learners with the development of flexibility and which learning opportunities require adaptation. Overall, the qualitative analysis aimed at determining, if students followed the hypothetical learning trajectory or deviated in particular parts. This was expected to provide helpful information, if and to what extent the intervention program matched the students' needs, and which parts of the program need to be refined in the future. Therefore, two questions were focused:

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

It was assumed that students would make use of the provided learning opportunities in different ways. Investigating such differences aimed at discovering typical patterns and systematic obstacles when developing the pursued flexibility. This may highlight potential "key processes", which require special attention when supporting students during the learning trajectory.

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

Under consideration of potential key processes from Q1, it was investigated if and how the students developed an ability to flexibly deal with arithmetic situations, and how this ability changed during the intervention. It was expected that learners learned to make use of the Dynamization Strategy and the Inversion Strategy and to formulate more complex, more comprehensive descriptions of arithmetic situations in the course of the intervention. Moreover, the analysis intended to investigate, if specific aspects, such as dealing with compare situations, would be more difficult to develop for some students.

5.4.2 Design and method

5.4.2.1 Context and case sampling

Before the intervention sessions started, four of the sixty students were selected from the experimental group based on pretest data. Since students with lower language skills struggle more with word-problem solving (e.g., Vilenius-Tuohimaa et al., 2008), it was of special interest, how such learners would make use of the learning opportunities to develop flexibility. Thus, pairs of students with lower scores in the reading comprehension test ELFE II were selected. The four selected students came from two different intervention groups, which were instructed by the same tutor (Group 1: Toni[15], Kim; Group 2: Alex, Chris; Tab. 5).

Subject	**ELFE II*** (0-100)	**Basic arithmetic skills and knowledge**** (0-1)	**Family language**
Toni	9.70	0.50	Croatian, German
Kim	21.20	0.38	Serbian, German
Alex	8.10	0.25	German, Italian
Chris	21.20	0.50	German

Tab. 5: Student profiles (*percentile rank with respect to the norm sample of the reading comprehension test ELFE II; **percentage of correct answers)

Their basic arithmetic skills and knowledge were at the lower (Kim, Alex) and lower average level (Toni, Chris). While Toni and Kim predominantly speak a language different from the instruction language at home, Alex and Chris speak mostly or exclusively German at home. Due to illness, Chris missed Sessions 5 and 6.

5.4.2.2 Overview of students' development

In the following, a rough overview of the development of the entire intervention sample (N = 60) and the four selected students is provided. Sessions 3, 6, and 8 included Verifying worksheets, which were linked pairwise by common tasks (see also Chapter 5.2.5). Students worked on these worksheets individually, and then discussed their ideas jointly. Their initial responses were scored dichotomously, and linked performance scores were calculated for each student and session using the 1PL Rasch model (Rasch, 1960)[16]. Fig. 21 displays the sample mean and the sample mean plus/minus one standard deviation of students' performance scores by

[15] The students' names were changed for data protection.
[16] The scale was anchored by setting the latent mean of person scores over all sessions to 0.

session (solid lines). A repeated measures ANOVA indicated a significant average progress of the intervention sample over the three sessions ($F(112.92, 2) = 18.94$, $p < .001$, $\eta^2 = .21$).

Fig. 21 also contains the performance scores of the four selected students. The development of students' performance scores differed substantially, some (e.g., Alex) developing roughly parallel to the sample mean, others showing substantial progress (e.g., Chris) or even a slight decrease in performance (e.g., Kim). The standard errors of individual performance estimates were between .57 and 1.60, showing that a reliable quantitative analysis of individual students' development in Verifying is not possible.

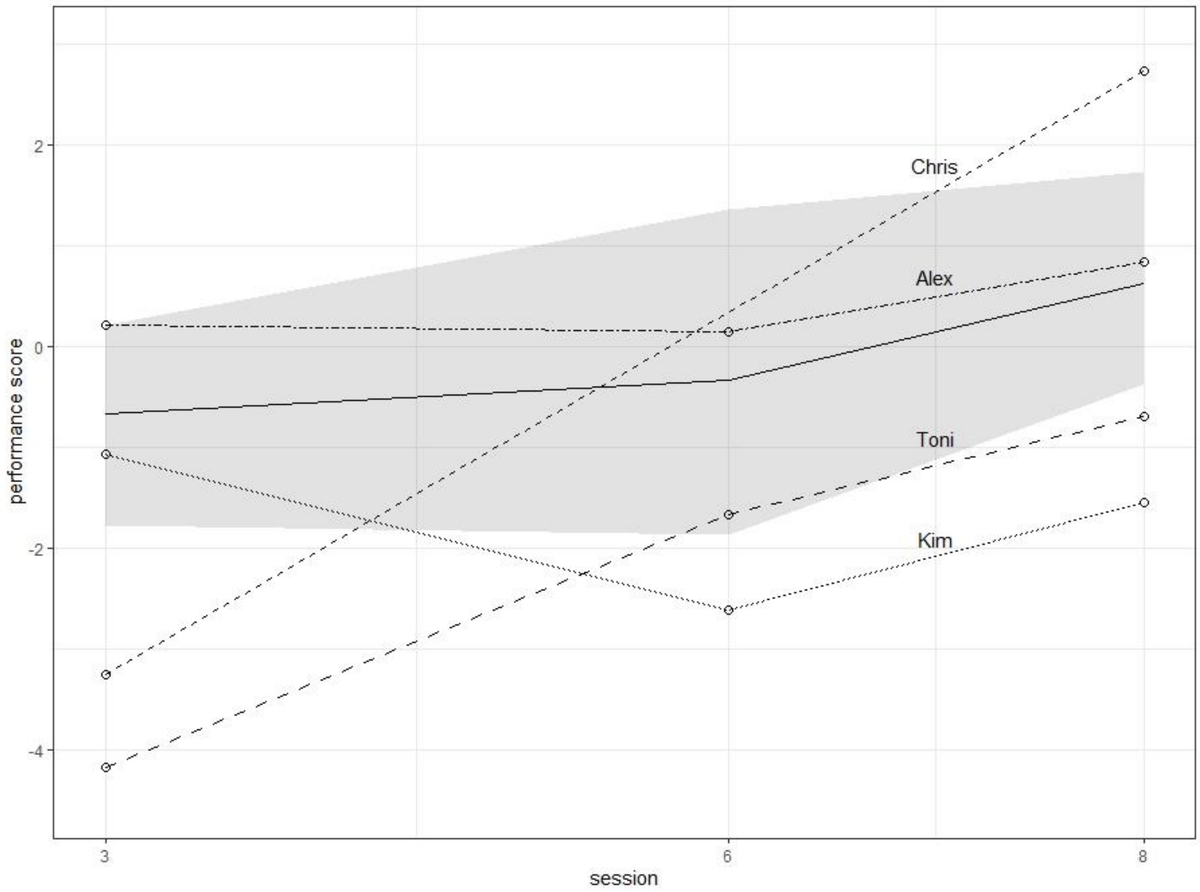

Fig. 21: Students' development in Verifying over Sessions 3, 6, and 8 (solid line: intervention sample mean; shaded area between 20% and 80% quantile)

5.4.2.3 Method: Qualitative content analysis

Consequently, a qualitative analysis was conducted to gain further insight into the students' development. For the qualitative analysis, all intervention sessions were recorded and transcribed. These transcripts and the students' work sheets were investigated following the principles of qualitative content analysis (Mayring, 2014). This category-based approach is characterized by a strong orientation on guiding research questions (Mayring, 2014). Each single student statement counted as one coding unit. Phases of group work were omitted, since contributions could not be attributed to single students. At first, a theory-based coding manual to identify different manifestations of flexibility was developed and then adapted during the analysis. Tab. 6 displays the final coding manual. The first two categories address, if students named statements on comparison and equalizing. Such statements were considered important prerequisites for developing flexibility (Chapter 5.2.4.2). Due to the specific characteristics of compare situations, additional subcategories were included here. The third and fourth category reflect the two strategies (Chapter 2.5.2). Whenever a student formulated a comparison or equalizing statement, and *then* immediately applied the Dynamization Strategy or Inversion Strategy by formulating a statement with a different semantic structure or a/s wording, the answer was coded as Category 3 or 4.

Category	Code	Sample statement
1. Verbalizing comparison		
1.1. Qualitative comparison	COM-QUAL	*"Susi has more marbles"*
1.2. Quantitative comparison		
1.2.1. Concrete set	CON	*"Susi has 7 marbles"*
1.2.2. 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		
3.1. 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"*
3.2. 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 perspective on mathematical relations		
4.1. Compare situations	COM-INV	"Susi has 4 marbles more than Max" → *"Max has 4 marbles less than Susi"*

4.2. 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.”*

Tab. 6: Coding manual

Each statement by a student on a single task was coded separately. For each statement, it was also coded, at which answer the respective task aimed. For open questions with more than one possible answer type (e.g., “What do you know about the situation?”), an extra code was used. The entire coding was conducted by two independent raters. Results indicate a very good interrater reliability with κ = 0.85 (Landis & Koch, 1977).

The development over sessions was analyzed for each student, and these paths of the different students were contrasted. This was done repeatedly with different emphases (Fig. 22) arising from the background and the design of the intervention. The analyses started from the coded data and proceeded to the raw data to check and enrich initial interpretations. In addition, remarkable observations in the raw data were checked for patterns in the codes to reconsider each student’s development.

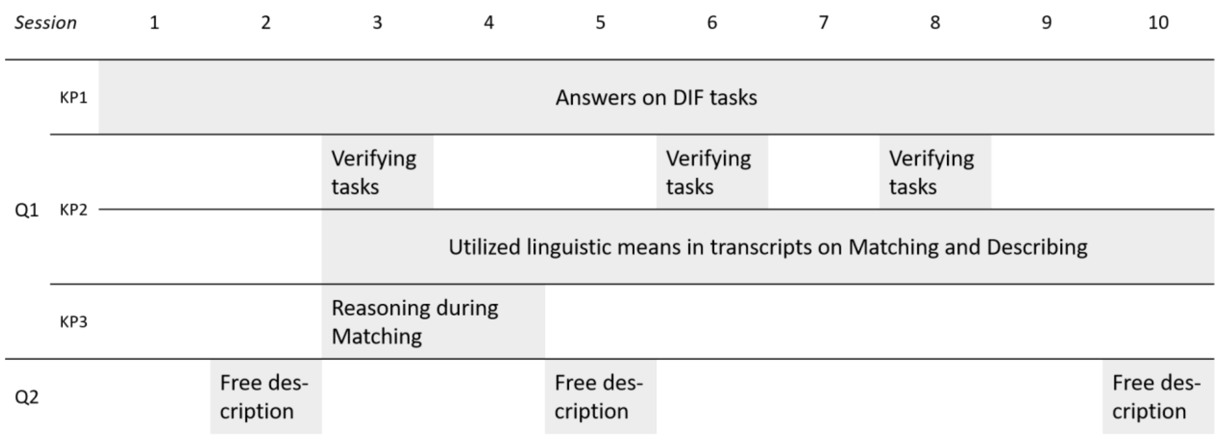

Fig. 22: Overview of different emphases during the data analysis, subdivided by research question (Q1, Q2) and emerging key processes (KP1, KP2, KP3)

5.4.3 Results

5.4.3.1 Q1: Uncovering potential key processes

Three major differences in students’ learning paths (in the following, these will be called *key processes*, “KP”) emerged during the analyses, which may point at parts of the learning trajectory, which were not sufficiently adapted to individual students yet.

(KP1) Distinguishing concrete sets and difference sets. One major difference in the students’ learning paths showed in the ways they interpreted difference sets in relational statements. To investigate these differences, students' answers on tasks that aimed at determining difference sets were considered (Fig. 22). Mostly, these answers related either to difference sets [DIF] or to concrete sets [CON]. Tab. 7 compares for each session, how often each student named a difference set or a concrete set when answering such tasks.

Alex and Chris mostly referred to difference sets and seemed to interpret relational statements correctly from the start (Tab. 7). Toni and Kim, however, mentioned concrete sets frequently and throughout the entire program, although tasks asked for difference sets. This indicates that they did not benefit fully from the corresponding learning opportunities in the Basics phase. During the program, they seemed to make some, but slower progress in this aspect.

	Toni		**Kim**		**Alex**		**Chris**	
Session	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

Tab. 7: Sets mentioned by students during tasks that required determining a difference set
(DIF = difference set, CON = concrete set)

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]. In this vein, 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 Kim worked on a statement such as "There are 7 sheep more than cows", Kim classified this statement as correct. Her answer indicates that she interpreted the qualitative relation 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 this excerpt:

Session 5, group 6

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

Kim: Well, there are less apples [COM-QUAL] and more bananas [COM-INV] in the bowl.

Tutor: Okay, can you express that more precisely?

Kim: In the bowl, there are… 3 apples…

Toni: I know it, too!

Kim: … 5 bananas… [CON]

After Kim 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 be a further sign that Kim linked numbers rather to concrete sets and relational statements separately to qualitative comparison. Similarly, Toni referred to concrete sets in relational statements systematically when answering a worksheet on the game *Hamstern* in Session 1, although the tutor had supported her before with interpreting the difference set by contrasting the sets verbally (design principle P6, Erath et al., 2021). 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 Toni to quantify the difference.

Tutor: Toni, what do you think, how many do I have more [DIF]?

Toni: You have 6 [CON].

Tutor: I do have 6, but how many do I have *more* than you? Think about it.

Toni: 4 [DIF].

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

Toni: 4 [EQ].

In Session 2, she distinguished concrete sets and difference sets correctly during a similar worksheet. However, Toni still seemed to struggle with this distinction occasionally until the end of the program (see Tab. 7).

Although difficulties with understanding difference sets were anticipated and thus considered in the learning trajectory, it was not expected that they appeared as systematically and frequently as it occurred with Toni and Kim. Thus, the difficulties could only partially be tackled in the Basics phase and were not resolved fully until the end of the intervention. It is plausible that such difficulties in the beginning might limit students' chance to profit from further parts of the program.

(KP2) Transferring from Verifying to Matching and Describing situations. Dealing successfully with Verifying tasks was assumed to stimulate flexible descriptions. In Sessions 3, 6, and 8, it was observed if students already developed initial flexibility. Tab. 8 shows how the four selected learners progressed differently on the three word problem types. Alex only gave a few, unsystematic incorrect answers throughout the three sessions. Since Chris misread the given situation with unknown difference set in Session 3 (Tab. 8), he answered almost all respective tasks incorrectly. 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, Chris answered all items correctly.

	Session 3		Session 6			Session 8		
Given situation	**DIF**	**DIF**	**DIF**	**DIF**	**COM**	**DIF**	**COM**	**REF**
Representation	Picture	Text	Picture	Text	Text	Text	Text	Text
Toni	4/6	5/6	6/6	4/6	3/6	5/6	6/6	1/6
Kim	6/6	5/6	5/6	4/6	1/6	4/6	4/6	1/6
Alex	6/6	5/6	6/6	6/6	5/6	6/6	5/6	5/6
Chris	6/6	1/6		*absent*		6/6	6/6	6/6

Tab. 8: 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)

Toni and Kim developed Verifying skills later and did not reach the same level as Alex and Chris. Situations involving two concrete sets (DIF) were easier for Toni and Kim than other types (Tab. 8). When verifying statements on situations with unknown compare set (COM, Session 6), both showed insecurities initially. Especially Kim seemed to struggle interpreting the qualitative relation of sets (Who has more?) of the given situation. In Session 8, both students showed substantial increase, 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 predicted in the learning trajectory, the students' ability to verify statements advanced from situations occurring in simpler word problem types (e.g., situations with unknown difference set) to more difficult ones (e.g., situations with unknown reference set) over the whole intervention. However, Toni and Kim progressed more slowly on tasks presenting relational sets verbally (situations with unknown compare/reference set). Progress in Verifying was intended to prepare students for Matching and Describing tasks by providing linguistic means and encouraging learners to use language for organizing their conceptual knowledge (Chapter 5.2.4).

Since macro-scaffolding was intended to support exactly this transfer from comprehending to actively describing situations in various ways (receptive vs. productive flexibility), it was analyzed how students made use of the provided language support in this transition more closely. Transcripts on Matching and Describing tasks were investigated with a specific focus on instances, in which macro-scaffolding (e.g., word cards, sentence frames, sentence starters) was integrated (Fig. 22). Chris and Alex had few problems to describe arithmetic situations quite early in the program. For Toni and Kim, the tutor offered more language support to formulate suitable statements and faded out the support more slowly. Toni had problems to formulate 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 is.
Tutor:	What should she do? Toni. Do you remember what we did there?
Toni:	If I one… eh? From Sebastian?
Tutor:	So, try to think about it again.
Toni:	If I one, then… at this tower… as tall as yours.
Tutor:	If you do what? "Then my tower is as tall as yours."
Toni:	If I…
Tutor:	What can you do, so that the tower is as tall as this one?
Toni:	One away?

Tutor: Exactly! Let's do that.

Toni: If I one… away… if I… eh?

Elisa: I know! If I one, then…

Tutor: You need more words.

Elisa: If I one brick… then…

Tutor: What do you do with the brick?

Elisa: If I take one brick away, then my tower is as tall as yours is.

Tutor: Fine, do that, Elisa, and now let's check if it's true… is the tower as tall as hers now?

Students: Yes!

Tutor: Okay, let's put the brick back.

[The next sentence frame is provided]

Toni: I know!

Tutor: Okay, Toni, you can try, you already started so well before.

Toni: 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, Toni managed to describe an equalization in the next task. From then on, her vocabulary on action verbs expanded continually. This indicates that the transfer from Verifying to more advanced parts of the hypothetical learning trajectory cannot be taken for granted. Individual language support seemed to be particularly important for Toni and Kim when developing flexibility. In this case, combining the sentence frame together with the support by her peer allowed Toni to progress on describing equalization.

(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, answers were contrasted with the codes [CON] and [DIF] on such tasks (Fig. 22). While Alex and Chris reasoned with statements on difference sets frequently (Alex: 4x DIF, Chris: 5x DIF), Toni and Kim mostly used the given concrete sets (Toni: 8x CON, Kim: 2x CON, 2x EQ). The following excerpts dealing with the question "What is the difference between the two pictures?" illustrate these observations:

Session 3, group 6

Toni: In this picture, Max has 2 [CON], and Susi has 4 [CON].

Tutor: Exactly.

Toni: And here, Max has 4 pieces [CON], and Susi has 2 cards [CON].

Tutor: Exactly.

Toni: Maybe, because they exchanged their cards?

It seems that, at this point, Toni did not use relations to contrast the situations. This complies with her tendency to 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. Kim matched an equalizing statement on the same situation to the correct picture later, but also referred only to concrete sets in her explanation. Alex used relational statements that involved a difference set to contrast the situations, and Chris even applied the Inversion Strategy.

Session 3, group 5

Alex: Here, Max has 2 more [DIF], and there, Susi has 2 more [DIF].

Tutor: Exactly. Can you say that in other words, Chris?

Chris: Here, Susi has 2 less [COM-INV], and there, Max has 2 less [COM-INV].

It seems that enhancing such discourse practices (design principle P1, Erath et al., 2021) can unveil students' conceptual knowledge and how they perceive situation structures. Here, providing useful linguistic means for alternative descriptions (e.g., "more") could have triggered Toni and Kim to focus on (qualitative and quantitative) relations and strengthened their awareness of difference sets. 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.

5.4.3.2 Q2: Development of flexibility

Under consideration of these key processes, the students' overall development of flexibility was examined. Three tasks on actively describing situations were selected to provide insights into students' development (Fig. 22). 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 intended to reveal, if they formulated varying descriptions and used language cognitively to enrich their situation models.

Tab. 9 lists the coding for each student. Overall, it becomes apparent that all four students produced more comprehensive descriptions and started to use at least one of the two strategies in the course of the intervention. This indicates that they could all enhance their ability to describe arithmetic situations flexibly to some extent.

Session	**Toni**	**Kim**	**Alex**	**Chris**
2	CON	CON	CON DIF	CON
5	<u>CON</u>	EQ EQ-INV	EQ EQ-INV	*absent*
10	CON <u>DIF</u> EQ EQ-INV	CON <u>COM-QUAL</u> DIF <u>COM-INV</u>	COM-QUAL DIF COM-INV COM-DYN EQ-INV	DIF COM-INV COM-DYN EQ EQ-INV

Tab. 9: 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). The underlined codes point to statements that did not match the given situation

Of the four students, all with lower language skills, different developmental patterns were observed. Alex was the only one to formulate statements on difference sets spontaneously already in Session 2. He may have the tendency to spontaneously focus on relations (“SFOR”, McMullen et al., 2013), which may draw his attention to relations between two sets. In spite of very low language skills and comparably low arithmetic pretest scores, he quickly adopted the two strategies. In Session 5, he added descriptions on equalization and inversion systematically, and also on dynamization in Session 10 (Tab. 9). In line with a consistent focus on relations, he preferred formulating relational statements. The following excerpt illustrates Alex’s systematic approach of describing situations flexibly:

Session 7, group 5

Tutor: Let’s have a look at this picture…

Alex: Can I start?

Tutor: …and describe what we see.

Alex: Can I start?

Tutor: Alex. Their names were Susi and Max.

Alex: Max has 4 keys [CON], Susi has 2 keys [CON].

Tutor: Please show that to me at the *Rechenschiffchen*. Susi is red *[color of chips]*. […] So, and now…

Alex: 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…

Tutor: As Max's.

Alex: …as Max's [COM-DYN]. If Max puts 2 keys away, then they're as many as Susi [EQ-INV].

Starting with higher language and arithmetic pretest scores than Alex, Chris first focused on concrete sets in Session 2 (Tab. 9). Although he missed two sessions, Chris adopted both strategies and developed flexibility with a strong focus on equalizing statements until Session 10. He and Alex required little language support beyond what was offered by the sequencing from Verifying over Matching to Describing tasks and the corresponding macro-scaffolding (KP2). Moreover, both distinguished concrete sets and difference sets already in early sessions (KP1).

In contrast, Toni 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 relational statements, or more adaptive language support. This is most likely a reason for her slower progress on Verifying tasks beyond those with unknown difference set. Given the hypothetical learning trajectory's structure, problems at Verifying probably made it difficult for her to work on further tasks meaningfully, and also, 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 (Tab. 9). For Toni, an adaptive deviation from the learning trajectory might have been promising. Indeed, students' reasoning (KP3) seems to provide indications, if such adaptations are necessary.

Kim started out with higher language skills, but with slightly lower basic arithmetic skills and knowledge than Toni. Similarly, she linked numbers primarily to concrete sets at first (KP1). However, she made progress at this key process. Like Toni, she struggled with Verifying tasks beyond those involving two concrete sets (KP2), and also, 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 relational statements and their inversion in Session 10 (Tab. 9). However, she then applied the Inversion Strategy incorrectly: After writing a correct statement, such as “There are more nuts and less tangerines [COM-QUAL, COM-INV]” she also wrote down the opposite of the given situation: “There are more tangerines and less nuts [COM-QUAL, COM-INV]” It seemed like Kim did apply inversion, but did not focus on describing the *same* situation. Also, a slower initial progress might have hampered her from benefiting from subsequent activities that intended to prepare learners for developing such strategies.

The findings from the qualitative analysis can be enriched by the students’ scores in the flexibility test and the word problem test (see Tab. 10). For Toni and Kim, the scores support the assumption that the two students did not fully benefit from the intervention as intended, except for Kim’s performance growth in the word problem test toward the third measurement. Chris barely seemed to benefit from the intervention in the two scales, probably since he was already on a high level in both tests. However, Alex’s performance increased substantially, especially in the word problem test. Further analyses should investigate, if the students’ performance increased for certain items in particular.

	Measurement	**Flexibility test***	**Word problem test****	
		(0-1)	(*M*)	(*SE*)
Toni	m1	0.80	-0.07	0.78
	m2	0.60	-1.89	0.72
	m3	0.65	-1.89	0.70
Kim	m1	0.50	-2.05	0.69
	m2	*absent*	*absent*	*absent*
	m3	0.35	-0.74	0.71
Alex	m1	0.65	-1.02	0.69
	m2	0.90	-0.06	0.79
	m3	0.85	2.03	1.54
Chris	m1	0.95	1.86	1.32
	m2	1.00	1.50	1.42
	m3	1.00	1.50	1.42

Tab. 10: Students’ scores in the flexibility test and the word problem test (*percentage of correct answers; **WLE ability parameters; m1 = pretest, m2 = posttest, m3 = follow-up test)

5.4.4 Discussion

The goal of the qualitative analysis was to gain detailed insights into the students’ development of flexibility (goal 2.5). Three potential key processes for successfully developing flexibility in dealing with arithmetic situations emerged from the analysis. From these observations, conclusions will be drawn on parts of the hypothetical learning trajectory, which 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 to develop flexibility. This points to a very specific interpretation of relational statements, but it shows how important developing a sound understanding of linguistic means is to describe situations from varying 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 to build up common understanding. In particular, this could be addressed with Verifying tasks contrasting situations that differentiate between statements such as “3 sheep more than cows” and “3 sheep, and more sheep than cows” in future revisions of the intervention.

KP2: As expected, transferring linguistic means from Verifying to Matching and Describing tasks was feasible. It is assumed that not only encountering linguistic means in the Verifying phase, but also discussing the use of these linguistic means is crucial. This is supported by the observation that language support by the tutor was vital for the transition to Describing. This underpins how important explicit support can be to help students to make use of linguistic means when reflecting on situation structures (Chapter 2.5.3.2). The observation that some learners needed the tutor’s language support (e.g., by asking to clarify or by reformulating the students’ answer) matches findings from other studies that identified conditions for successfully enhancing language (Erath et al., 2021; Pöhler & Prediger, 2015; Prediger & Wessel, 2013).

KP3: The observations reveal the power of using language communicatively to uncover students’ conceptual knowledge and flexibility regarding situation structures. If learners focus on specific semantic structures or on concrete sets in their descriptions, this might indicate that other situation structures should be discussed more intensively with the student. 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 consequently, to get easier access to both strategies.

Secondly, the analysis gave information on the students’ development of flexibility (Q2) and consequently, students’ cognitive use of language during word-problem solving. Some students progressed mostly as predicted. Others made progress along the hypothetical learning trajectory, but took substantially more time. This supports assumptions that the intervention is a feasible way to foster at least some students’ flexibility (see Chapter 3.2.2). 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 stronger in compare situations. However, the fact that Kim and Toni made some progress shows again that initial problems do not necessarily imply that learners cannot develop flexibility during the intervention. The differences rather seemed to derive from a primarily qualitative interpretation of relational statements. This qualitative interpretation indicates that Toni and Kim already identified relations between quantities, but without referring to number words, which is allocated to earlier levels of number concept acquisition (Krajewski & Schneider, 2009). Linking relations with number words, which constitutes the highest level of competence in models on number concept acquisition (e.g., Krajewski, 2008), seems to be a prerequisite not only for successfully solving compare problems, but also for developing flexibility in dealing with arithmetic situations. Future research needs to investigate, if flexibility develops in parallel or subsequent to number concept acquisition.

Also other factors such as prior knowledge may be the basis of different learning paths. Alex developed substantial flexibility despite lower language skills and basic arithmetic skills and knowledge, which corroborates the finding from the quantiative analysis that the intervention is also suitable for supporting learners with lower language skills (see Chapter 5.3.3, Chapter 5.3.4.2). It seems that Alex already had a tendency to focus on quantitative relations at the beginning (McMullen et al., 2013), which might have given him a good starting point to make use of the two strategies. Future research may investigate reasons why students' learning paths varied as observed by taking into account individual and didactial aspects more systematically.

Although qualitative analysis can uncover relevant aspects that were hidden from a more summative, quantitative approach, the analysis has to be viewed in the light of some limitations. For example, it cannot be ruled out that affective factors, such as the students' motivation or their self-efficacy (Verschaffel et al., 2015), influenced the four students' development. The analysis is restricted to four pre-selected students. However, by repeatedly sampling transcripts and contrasting their cases, the conclusions were put to a first test. Further analysis need to explore in more detail and more systematically, why, when, and to what extent the reported key processes occur in a larger sample, and also, which role these key processes play in developing flexibility. Moreover, the newly created learning trajectory requires further modification (Simon, 1995). More adaptive learning opportunities will be necessary to better meet the learners' needs when building conceptual knowledge of relations as a quantitative phenomenon. However, within this study, the predetermined structure of the intervention program helped to contrast learners' paths reliably.

In spite of these limitations, the analyses show that fostering the pursued flexibility is generally possible and endorse the feasibility of the chosen approach to support students in constructing richer, more accurate situation models. The results provide a starting point to address students' difficulties with word problems by enhancing their flexibility in dealing with arithmetic situations. Encouraging the cognitive and communicative use of language and integrating the design principles by Erath et al. (2021) has turned out to be an expedient approach to achieving this goal. The hypothetical learning trajectory proved to correspond roughly to the students' learning paths. However, as expected, the development of flexibility goes along with substantial and possibly systematic heterogeneity. Therefore, the learning trajectory and the intervention program needs to be refined in future research to better meet the needs of heterogeneous groups.

6 General discussion

6.1 Contribution of this work to the field

The main goal of this project was to investigate the construct of flexibility in dealing with arithmetic situations. In preparation, the long-standing field of additive one-step word problems and their difficulty was revisited. The preliminary study provided updated information on the difficulty of additive one-step word problems in grade 2 (goal 1.1). In contrast to prior findings (e.g., Carpenter et al., 1981; Stern, 1998; Verschaffel & De Corte, 1997), the results indicated higher and more homogeneous solution rates depending on the semantic structure. Consequently, it could not be replicated that compare problems are particularly difficult compared to other semantic structures. What seemed more important was, if the a/s wording and the directly applicable operation of a word problem were consistent or inconsistent. The particular difficulty of inconsistent word problems was conform to the findings reported in prior studies (e.g., Lewis & Mayer, 1987; Verschaffel, De Corte, & Pauwels, 1992) and again replicated within the intervention study. These updated findings may work as a foundation for future research. Based on the presented findings, intervention studies that enhance the solution of additive one-step word problems should foreground understanding inconsistent word problems across all semantic structures instead of solely focusing on compare problems. Moreover, a preference for additive over subtractive mathematical models was observed in the preliminary study. Supporting learners with understanding the symmetry of relations (Stern, 1993) and actions may not only be a way to tackle inconsistent word problems, but also to reduce preference effects by emphasizing addition and subtraction as complementary operations (Renkl & Stern, 1994). The preliminary study also systematized prior results on the difficulty of additive one-step word problems (1.2). By systematically varying the situation structure and the context, the preliminary study confirmed the hypothesis that the context (the involved subjects, objects, and numbers) of a word problem only had a minimal effect on the students' solution rates. Consequently, future investigations on additive one-step word problems can neglect the influence of context features (given comparable conditions).

Based on these findings, the construct of flexibility in dealing with arithmetic situations was investigated. This investigation made a contribution to the field on different levels: [1] On a theoretical level, a newly developed construct was conceptualized based on suggestions in the literature (Greeno, 1980; Stern, 1993). A special characteristic of this construct is that it is based on the level of the situation structure and thus gets by without conducting mathematical operations. [2] On a methodical level, the project provided a starting point how this flexibility can be operationalized. The corresponding test instrument was tested within both studies and turned out as a reliable instrument to measure a certain manifestation of flexibility. Observations on this innovative test instrument can be used to refine the instrument in the future. [3] On a didactical level, the project generated a concept how learners could be supported with developing flexibility. This concept suggests a hypothetical learning trajectory (Simon, 1995) and corresponding activities for fostering this new ability construct. In the concept, ideas are put forward how conceptual learning and language learning can be intertwined (Pöhler & Prediger, 2015) to foster the pursued flexibility. This is implemented by considering current standards of design principles for enhancing language when learning mathematics (Erath et al., 2021). Also the role of using language cognitively and communicatively for learning mathematics, which is increasingly emphasized in recent literature (Haag et al., 2015; Kempert et al., 2019; Peng et al., 2020) as a major mechanism explaining language-related differences in mathematics performance, was considered when generating the concept. [4] On a level of empirical evidence, the project provided information, if and how learners make use of flexibility and benefited from the intervention program. One question was, if providing learners with structurally similar pairs of word problems was sufficient to stimulate learners to make use of flexibility (2.1). Indeed, there was hardly any evidence that learners recognized structurally similar situations, and spontaneously used this information to solve subsequent word problems. The observation that even a hint on the structural similarity did not stimulate students to make use of the similar features, makes it unlikely that this would merely be an effect of socio-mathematical norms in the classroom context (Yackel & Cobb, 1996).

Since flexibility could not be observed with this method in the preliminary study, it was investigated, if the pursued flexibility could be fostered with an explicit training. The intervention study served as a “feasibility study” that investigated, if flexibility in dealing with arithmetic situations could be fostered at all by the suggestions of Greeno (1980) and Stern (1993) (2.2). Indeed, the suggestions in literature proved effective. The quantitative as well as the qualitative analyses confirmed that flexibility can be developed and that the intervention is a feasible way to foster such flexibility. In particular, an increased receptive flexibility was observed in the qualitative (Verifying sheets) as well as in the quantitative analyses at least in tendency (flexibility test). Since this flexibility was expected to support learners with solving difficult word problems, it was also investigated, if students’ performance in word-problem solving increased after the intervention (2.3). The analyses revealed a significant performance growth in the experimental group for word-problem solving. This indicates that the intervention was effective to support learners on average with solving difficult word problems and thus adds a functioning intervention to the field. Moreover, this project shows 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 completely left aside.

Furthermore, it was investigated how language skills influenced the development of flexibility. Various scenarios seemed plausible: It could be that either students with lower or with higher language skills may benefit stronger (Kalyuga et al., 2003; Merton, 1968), but also equal benefits were assumed possible. Indeed, students benefitted comparably from the intervention regardless their language skills (2.4). This supports the findings by Prediger and Wessel (2018) that enhancing language in mathematics interventions can be helpful for learners with lower and higher language skills. Moreover, the intervention provides a language-sensitive approach to support learners with different levels of language skills with word-problem solving.

Finally, the analyses provided insight into the development of flexibility (2.5). Overall, the hypothetical learning trajectory appeared to be suitable for most students (Huang et al., 2019; Simon, 1995). It was feasible for learners to adopt linguistic means and use them for describing arithmetic situations flexibly. The program indeed elicited more comprehensive descriptions by encouraging using the Dynamization Strategy and the Inversion Strategy. As expected, receptive flexibility seemed to precede productive flexibility: Only when the sessions on receptive flexibility were completed (around Session 5, see Chapter 5.2.4.1), students' productive flexibility became more differentiated in the form of more comprehensive descriptions (see Tab. 9). As realized in the coding manual, being able to understand and express relational statements and the contained difference sets appeared to be a prerequisite for developing flexibility. A particularly important key process was to understand numbers not only as a description of concrete sets, but also as a way to describe the relation between two sets. Relational statements being a requisite for developing flexibility indicates that flexibility in dealing with arithmetic situations also evolves *after* completing higher levels of number concept acquisition (Fritz et al., 2018; Krajewski & Schneider, 2009).

6.2 Open questions

6.2.1 How to conceptualize and measure the construct of flexibility?

The analyses could provide first insights into the newly developed construct of flexibility in dealing with arithmetic situations. However, there are still unanswered questions. The construct of flexibility in dealing with arithmetic situations is based on applying the Dynamization Strategy and the Inversion Strategy. These strategies were derived from suggestions in the literature (Greeno, 1980; Stern, 1993) and include that students contrast features of the semantic structure and the a/s wording. Another feature of the situation structure that could be used to contrast different perspectives on an arithmetic situation, but has not been implemented in the intervention yet, is the unknown set. After first considerations, the unknown set does not seem suitable for reinterpreting word problems as in the Dynamization or Inversion Strategy. However, it seems possible that contrasting situations with different unknown sets may at least provide fruitful learning opportunities and further contribute to students' conceptual knowledge (Rittle-Johnson et al., 2001). This could be implemented by contrasting different perspectives on a given situation, in which only the unknown set is varied. For example, students could be asked, if the following descriptions describe the same situation:

- “Susi has 3 marbles. Max has 5 marbles.”
- “Susi has 3 marbles. Max has 2 marbles more than Susi.”
- “Susi has 3 marbles. She has 2 marbles less than Max.”
- “Max has 5 marbles. He has 2 marbles more than Susi.”
- “Max has 5 marbles. Susi has 2 marbles less than Max.”
- “Max has 5 marbles. Susi has 3 marbles.”

By contrasting language pieces regarding different unknown sets systematically, students may discover connections between the problem types and the involved sets (Erath et al., 2021; Kullberg et al., 2017). Future interventions may include this learning opportunity and investigate, if it is suitable to foster students’ flexibility in dealing with arithmetic situations.

Moreover, it is still an open question, how the construct of flexibility in dealing with arithmetic situations can be measured validly. Although the newly developed test instrument turned out as a reliable way to measure (at least one part of) flexibility, further adaptations are necessary. One observation was that the test was quite easy and thus did not differentiate well between high-achieving students’ performance. As stated, one explanation may be that the test instrument did only measure receptive, but not productive flexibility. However, productive flexibility may be important for word-problem solving, since developing productive flexibility can stimulate the cognitive use of language: When students learn to actively describe situations in various ways (and thus use language communicatively), this may also induce them to perform this activity mentally and make use of language during thinking. In other words, communication can be used in this case to create learning opportunities for using language cognitively and even reinforce the cognitive use (Maier & Schweiger, 1999). Adding a productive component to the test could be achieved by integrating task formats, in which students would describe a situation flexibly orally or in writing (“What do you know about the situation?”). However, it is unclear yet, how productive flexibility can be measured, in particular with this age group and with larger samples. Measuring flexibility orally may only work in the one-to-one context. Therefore, this format would be unsuitable for larger samples as in the intervention study. Describing various situations flexibly in writing may be viable for larger samples. However, generating comprehensive descriptions may be too time-consuming (Xu et al., 2017), as it was also observed in Session 10, in which this activity was actually implemented in written form with the experimental group. Especially second graders with lower language skills may struggle with writing longer texts. Therefore, measuring productive flexibility in writing may only be possible in limited form with younger students. Future research needs to investigate, if and how such instruments can be used to measure students’ flexibility in second grade, and moreover, find ways how to measure flexibility in dealing with arithmetic situations appropriately for different age groups.

Research has investigated the measurement of flexibility in other fields of mathematics (e.g., for linear equations in middle school, Xu et al., 2017). In this context, flexibility is separated into two components: *potential flexibility* (knowledge of strategies) and *practical flexibility* (use of strategies). There may be some parallels between these constructs and the separation into receptive and productive flexibility. Potential and receptive flexibility both rather refer to a competence related to strategies, while practical and productive flexibility seem to depend on the actual performance of strategies. Similar to the assumed relationship of receptive and productive flexibility, potential flexibility is assumed to precede practical flexibility (Xu et al., 2017). Therefore, it may be helpful to draw on the experiences made with measuring flexibility in this context. Xu et al. (2017) report that potential flexibility is often measured via processes of recognition and evaluation. Especially the evaluation process resembles the Verifying activity implemented in the intervention study, in which learners evaluate the validity of given statements. Practical flexibility is often measured via processes of generation (Xu et al., 2017). This, on the other hand, corresponds with actively describing situations flexibly. Xu et al. (2017) recommend to measure both the potential and the practical component of flexibility, since potential flexibility positively predicts practical flexibility (Liu et al., 2018) and implementing strategies may be more demanding for learners than only possessing the knowledge on such strategies. They hypothesize that relying on only one component may either lead to overestimating the students' abilities, if only assessing potential flexibility, or to underestimation, if only assessing practical flexibility. It is an open question, if this is also valid for the separation into receptive and productive flexibility.

Another open question that also relates to the measurement of flexibility refers to the items of the flexibility test. As stated in Chapter 5.2.2.1, the test contains items that are associated either with the Dynamization Strategy or with the Inversion Strategy. It would be interesting to separate the current test instrument into two scales and compare their difficulty – provided that this separation is psychometrically reasonable. Tab. 9 from the qualitative analyses indicates a preference of the four students for the Inversion Strategy at the end of the intervention. It is unclear, if this is related to the difficulty of the two strategies, or if this is solely an effect of putting the emphasis on inversion in the intervention (e.g., in the extra phase “Symmetry of relations”). The difficulty of both strategies needs to be investigated in future analyses, on the one side, in the context of the test instrument, and on the other side, within the scope of process data from the intervention.

6.2.2 Which other, not yet considered factors might be a prerequisite for developing flexibility?

The project has considered various factors, which may influence the development of flexibility, such as language skills or basic arithmetic skills and knowledge. In addition, the qualitative analysis has indicated that understanding relational statements and difference sets may be a prerequisite to develop flexibility. However, there may also be other factors that were not considered in the project yet, but play a role in the students' development. Two factors will be discussed in the following: (1) the role of working memory, and (2) the role of metacognition.

(1) It is not clear yet, which role working memory plays in dealing flexibly with arithmetic situations. It seems plausible that students with lower working memory capacity are also limited in reinterpreting the situation structure of a word problem. Enriching the individual situation model with further features may require too many cognitive resources (Stern, 1998), so that this strategy may only make it more challenging for learners, who already struggle because of the particular demands of word problems on working memory capacity (Wang et al., 2016). For dealing flexibly with arithmetic situations, one particular component of working memory may play a special role: The *phonological loop* may work as a temporary storage of verbal information (Fung & Swanson, 2017). The higher the capacity to store verbal information, for instance, various descriptions of a situation, the more extensively learners might be able to construct their situation model. This hypothesis needs to be tested in further investigations.

(2) Another factor influencing students' development of flexibility may be the availability of metacognitive or self-regulatory strategies. These are strategies that "learners apply prior, during, and after the execution of a cognitive task to regulate one's own thinking and learning" (Verschaffel et al., 2020, p. 5). During word-problem solving, students constantly need to monitor their situation model and check it for consistency and plausibility, in particular when new information is integrated as an inference (Stephany, 2021). Making use of flexibility in dealing with arithmetic situations is based on drawing upon such inferences by integrating further features of the situation structure into the situation model (see Chapter 2.5.1). Thus, it seems particularly important in this context to monitor, if the constructed model is adequate and likely to lead to a successful solution (Schoenfeld, 2016). In case students struggle during the solution process, they may need to decide *actively* to enrich their situation model with further features of the situation structure. In the current intervention program, no actual word problems were solved and the active recourse to further features of the situation structure was not practiced during word-problem solving. However, it is possible that students already used metacognitive or self-regulatory strategies more or less successfully during the word-problem tests. Future investigations could analyze students' monitoring of their solution process during word-problem solving to uncover, if students actively decided to make use of flexibility. This could, for example, be explored with the "thinking-aloud" method (Lewis, 1982). Comparing

students' metacognitive and self-regulatory processes before and after the intervention could provide information, if the intervention's activities were already sufficient to trigger such monitoring spontaneously. If this was not the case, studies could investigate, if students benefit from an explicit training, in addition to the intervention, in which students learn to monitor the consistency and the plausibility of their situation model and to decide consciously to make use of flexibility, if needed.

6.3 Implications for future research

The findings from the project provide a starting point for further analyses on the role of flexibility in dealing with arithmetic situations during word-problem solving. The intervention study controlled for basic arithmetic skills and knowledge, language skills, general cognitive abilities, and the students' personal background (e.g., socio-economic status). However, other variables may play a role in developing flexibility in dealing with arithmetic situations. In terms of domain-general abilities, only general cognitive abilities have been included as a control variable. As outlined in Chapter 2.3.2, also the students' working memory influences word-problem solving (Wang et al., 2016) and is likely to influence the development of the pursued flexibility (Chapter 6.2.2). Moreover, inhibitory control is central when students encounter difficult word problems (Verschaffel et al., 2020). Even if students possess sufficient flexibility to create an adequate situation model, it is possible that they may be misled by certain stimuli (Páchová & Vondrová, 2021) or consciously choose accustomed strategies over innovative strategies (Xu et al., 2017). Learners still need to inhibit such stimuli (e.g., key words) as an indicator for the underlying mathematical operation and consciously decide to apply the suggested strategies. Therefore, working memory and inhibitory control should be included as control variables in further studies.

Besides cognitive factors, also affective factors may play a role in developing flexibility (Verschaffel et al., 2015). The influence of self-efficacy, which refers to "people's beliefs in their capabilities" (Bandura, 1989, p. 730), has already been confirmed for word-problem solving in general (e.g., Pajares & Miller, 1994). Liu et al. (2018) investigated the role of self-efficacy in the context of potential and practical flexibility. Indeed, self-efficacy seems to moderate the relationship between these two flexibility types, suggesting that "potential flexibility may lead to different degrees of practical flexibility depending on different levels of beliefs" (Liu et al., 2018, p. 1). Although this can not necessarily be transferred directly to receptive and productive flexibility in dealing with arithmetic situations, it still seems plausible that students' beliefs may play a particular role in actively describing situations (Verschaffel et al., 2015). Future research needs to include measures of self-efficacy to clarify the role of this variable in developing flexibility. Furthermore, emotional factors, such as the students' motivation or engagement during the sessions (Verschaffel et al., 2015), may influence the students' development of flexibility. Despite collecting process data on the students' participation and engagement during the intervention (see Chapter 5.2.5), the role of motivation or engagement in developing flexibility with the suggested strategies needs to be explored more systematically in future research.

Findings from this project suggest that flexibility in dealing with arithmetic situations also evolves *after* completing higher levels of number concept acquisition (see Chapter 6.1). However, it needs to be investigated in more detail, how number concept acquisition (Fritz et al., 2018; Krajewski & Schneider, 2009) is related to this flexibility and how it affects the development of this ability construct. Prior research by Stern (1993) has revealed that the majority of first graders struggled with recognizing the equivalence of symmetrical relational statements. To investigate this, students were confronted with two symmetrical relational statements and asked to decide, if both, neither, or only one of these sentences were true. In contrast, the flexibility test used in this project explicitly asked second graders to decide, if the two relational statements match. It cannot be determined clearly yet, why these findings differ. The high solution rates in the flexibility test, which were already achieved in the pretest, and also achieved by the control group at later measurements, indicate that at least in some way, the students' flexibility may increase from first to second grade. Another plausible explanation may be that classroom instruction has changed, and Stern's findings do not apply fully to current instruction anymore. In addition, different methods were used: Stern (1993) only measured the ability to recognize inversion, while the flexibility test also included items on the Dynamization Strategy. Future research needs to align the findings from the dissertation project with prior findings. For this, studies should investigate systematically with the same methods across the grades, when flexibility begins to develop and how this flexibility advances in relation to number concept acquisition.

The intervention program itself was effective, but needs to be adapted and refined in terms of addressing a wider range of students with heterogeneous prerequisites and needs. In particular, further material for supporting students with understanding difference sets needs to be developed and tested for its effect. This material should emphasize the difference between concrete sets and difference sets, for example by systematically contrasting statements such as "I have three marbles" and "I have three marbles *more*" (design principle P6, Erath et al., 2021). The analyses also indicate that a stronger focus on relational statements and their linguistic features is necessary. Huang et al. (2019) conducted an effective intervention study on word-problem solving, in which the symmetrical relational statements (e.g., "A is x more than B.", "B is x less than A.") were combined with a third way to describe a difference set. This description eliminated the a/s wording: "The difference between A and B is x." Adding this perspective may strengthen the link between symmetrical relational statements and further illustrate the relationship between two concrete sets. This new perspective may not only clarify the intention of relational statements (expressing a difference between two sets) in greater detail, but also enrich the students' conceptual knowledge on the comparison of sets.

On a broader perspective, it would be interesting, if the construct of flexibility in dealing with arithmetic situations is transferable to mathematical subjects other than additive one-step word problems. First, it needs to be investigated, if the construct can be applied to more complex additive word problems, for example multi-step word problems or more authentic, real-world problems. Moreover, the idea of enriching situation models by linking perspectives on the situation may also be viable for multiplication and division. For example, learners could not only view a multiplicative situation as a temporal-successive process (e.g., "Susi went down to the basement 5 times and carried 3 bottles of water each time."), but also as a result of this process. The semantic structure could also be interpreted as a spatial-simultaneous state (e.g., "Now, the bottles of water may be arranged in 5 rows of 3 bottles each.") (Padberg & Büchter, 2015). Future research may investigate with similar methods used in the intervention program, if developing flexibility in dealing with arithmetic situations is also possible for multiplication and division.

6.4 Implications for teaching

The research on fostering students how to develop flexibility in dealing with arithmetic situations is still at an early stage. Nonetheless, it is possible to derive implications for teaching from this dissertation project. In general, the activities in the intervention program and integration of Dynamization Strategy and Inversion Strategy have proven effective and the intervention design made it possible to confirm the feasibility of the chosen approach. However, since the students were tutored in addition to regular mathematics lessons, it is not possible to draw conclusions for normal circumstances. In the future, the program needs to be integrated into regular mathematics lessons and compared with conventional instruction on word-problem solving. Since there is only limited space in mathematics lessons, the activities of the intervention program need to be transformed into practicable formats for classroom instruction. One challenge in implementing this idea will be to condense the ten sessions of the intervention program without losing efficacy. For this, the learning trajectory and the design principles for enhancing language when learning mathematics should be maintained as good as possible.

The results from the analyses also provide information, which focus may enhance the quality of mathematics instruction. The newly gained insights may be integrated in the tasks in students' textbooks or also in teacher education and training. As outlined in Chapter 3.1, learning opportunities in textbooks influence students' performance in mathematics. An analysis confirmed for solving compare problems in particular that a textbook's topic-specific quality influences students' achievement in this content area (Sievert et al., 2021). Since textbooks influence, which learning opportunities are provided (Van den Ham & Heinze, 2018), refining textbooks can work as an impulse to advance mathematics instruction. The analyses from this dissertation project indicated that the unknown set and the consistency of a word problem strongly influenced its difficulty. Further, developing the pursued flexibility seems to require that students systematically encounter and contrast various features of the situation structure, so that they can draw on these features when enriching their situation model. However, the current distribution of additive one-step word problems and their subtypes seems to be rather imbalanced (Von Damnitz, 2020). In particular, word problems with unknown start set, reference set, or subset, and also certain inconsistent word problems occurred only sporadically in this textbook analysis. Consequently, textbooks should be enriched with more diverse learning opportunities on different subtypes of word problems. The effectiveness of the intervention study suggests that such learning opportunities do not need to be exclusively based on working with mathematical operations, but can also take place on the level of the situation structure. For example, tasks that contrast different features of the situation structure may be a valuable complement to current textbooks. The findings from this dissertation project support implications by Stern (1993), who emphasized the importance of understanding the symmetry of relations, and adds the observation that also understanding the symmetry of changes is crucial for dealing flexibly with arithmetic situations. It may be helpful to integrate tasks into textbooks that illustrate this symmetry and provide necessary linguistic means for expressing different perspectives on arithmetic situations.

Also in teacher education and training, minor refinements could lay the groundwork for a mathematics instruction in the context of additive one-step word problems, which enhances language and conceptual knowledge by interconnecting different perspectives on and descriptions of arithmetic situations. This begins with addressing uncommon or difficult word problem types in teacher education. Prospective teachers could not only learn to differentiate different word problem types, but also to generate them on their own. Moreover, teacher education and training could provide teachers with ideas how they can foster students' flexibility in dealing with arithmetic situations and why this is desirable. Sensitizing teachers for challenges in dealing with additive one-step word problems and illustrating such empirically effective instructional approaches may improve the quality of teacher education, and with that, also mathematics instruction.

7 References

Abedi, J., Courtney, M., Leon, S., Kao, J., & Azzam, T. (2006). *English language learners and math achievement: A study of opportunity to learn and language accomodation*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, University of California.

Artemenko, C., Pixner, S., Moeller, K., & Nuerk, H.-C. (2018). Longitudinal development of subtraction performance in elementary school. *British Journal of Developmental Psychology, 36*(2), 188–205.

Awisati, F., & González-Sancho, C. (2016). *PISA 2015 Ergebnisse: Exzellenz und Chancengerechtigkeit in der Bildung. Bd. I*. Bielefeld: W. Bertelsmann.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology, 25*(5), 729–735.

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.

Barwell, R. (2005). Ambiguity in the mathematics classroom. *Language and Education, 19*(2), 117–125.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(7), 1–23.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., . . . Weiß, M. (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.

Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 323–410). Opladen: Leske + Budrich.

Beal, C. R., Adams, N. M., & Cohen, P. R. (2010). Reading proficiency and mathematics problem solving by high school English language learners. *Urban Education, 45*(1), 58–74.

Benz, C. (2005). *Erfolgsquoten, Rechenmethoden, Lösungswege und Fehler von Schülerinnen und Schülern bei Aufgaben zur Addition und Subtraktion im Zahlenraum bis 100*. Hildesheim: Franzbecker.

Bjork, I. M., & Bowyer-Crane, C. (2013). Cognitive skills used to solve mathematical word problems and numerical operations: A study of 6- to 7-year-old children. *European Journal of Psychology of Education, 28*(4), 1345–1360.

Blum, W., & Leiß, D. (2005). Modellieren im Unterricht mit der "Tanken"-Aufgabe. *Mathematik lehren, 128*, 18–46.

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). Chichester, UK: Horwood Publishing.

Bochnik, K. (2017). *Sprachbezogene Merkmale als Erklärung für Disparitäten mathematischer Leistung: Differenzierte Analysen im Rahmen einer Längsschnittstudie in der dritten Jahrgangsstufe*. Münster: Waxmann.

Bochnik, K., & Ufer, S. (2016). Die Rolle (fach-)sprachlicher Kompetenzen zur Erklärung mathematischer Kompetenzunterschiede zwischen Kindern mit deutscher und nicht-deutscher Familiensprache. *Zeitschrift für Grundschulforschung, 9*(1), 135–147.

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G., & Valtin, R. (2003). *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.

Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Ed.), *Soziale Ungleichheiten* (pp. 183–198). Göttingen: Schwartz.

Breidenbach, W. (1969). *Methodik des Mathematikunterrichts in Grund- und Hauptschulen*. Hannover: Schroedel.

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

Cankoy, O., & Özder, H. (2011). The influence of visual representations and context on mathematical word problem solving. *Pamukkale University Journal of Education, 2*(30), 91–100.

Carpenter, T. P., Hiebert, J., & Moser, J. M. (1981). Problem structure and first-grade children's initial solution processes for simple addition and subtraction problems. *Journal for Research in Mathematics Education, 12*(1), 27–39.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1–22.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam: Elsevier.

Civil, M. (2008). Language and mathematics: Immigrant parents' participation in school. In O. Figueras, J. L. Cortina, S. Alatorre, T. Rojano, & A. Sepúlveda (Eds.), *Proceedings of the Joint Meeting of the 32nd Conference of the International Group for the Psychology of Mathematics Education and the XX North American Chapter* (Vol. 2, pp. 329–336). Morelia, Mexico: PME.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Coley, R. J. (2002). *An uneven start: Indicators of inequality in school readiness. Policy information report*. Princeton, NJ: Educational Testing Service, Policy Information Center.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20*(4), 405–438.

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. *Encyclopedia of Language and Education, 2*(2), 71–83.

Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology, 137*, 57–75.

Czocher, J. A. (2018). How does validating activity contribute to the modeling process? *Educational Studies in Mathematics, 99*(2), 137–159.

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.

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.

Davis-Dorsey, J., Ross, S. M., & Morrison, G. R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology, 83*(1), 61–68.

De Corte, E., & Verschaffel, L. (1987). The effects of semantic and non-semantic factors on young children's solutions of elementary addition and subtraction word problems. In J. C. Bergeron, N. Herscovics, & C. Kieran (Eds.), *Proceedings of the 11th Conference of the International Group for the Psychology of Mathematics Education* (pp. 855–861). Montreal: PME.

Deschuyteneer, M., De Rammelaere, S., & Fias, W. (2005). The addition of two-digit numbers: Exploring carry versus no-carry problems. *Psychology Science, 47*(1), 74–83.

Dröse, J. (2019). *Textaufgaben lesen und verstehen lernen: Entwicklungsforschungsstudie zur mathematikspezifischen Leseverständnisförderung*. Wiesbaden: Springer.

Dröse, J., & Prediger, S. (2019). Enhancing fifth graders' awareness of syntactic features in mathematical word problems: A design research study on the variation principle. *Journal für Mathematik-Didaktik*, 1–32.

Ehmke, T., Hohensee, F., Siegle, T., & Prenzel, M. (2006). Soziale Herkunft, elterliche Unterstützungsprozesse und Kompetenzentwicklung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost, & U. Schiefele (Eds.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (pp. 225–248). Münster: Waxmann.

Ehmke, T., & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider, & P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 231–254). Münster: Waxmann.

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.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.

Franke, M., & Ruwisch, S. (2010). *Didaktik des Sachrechnens in der Grundschule*. Heidelberg: Springer.

Fritz, A., Ehlert, A., & Leutner, D. (2018). Arithmetische Konzepte aus kognitiv-entwicklungspsychologischer Sicht. *Journal für Mathematik-Didaktik, 39*(1), 7–41.

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

Fung, W., & Swanson, H. L. (2017). Working memory components that predict word problem solving: Is it merely a function of reading, calculation, and fluid intelligence? *Memory & Cognition, 45*(5), 804–823.

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.

Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., . . . Fennema, E. (1997). Children's conceptual structures for multidigit numbers and methods of multidigit addition and subtraction. *Journal for Research in Mathematics Education, 28*(2), 130–162.

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.

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.

García, O. (2017). Critical multilingual language awareness and teacher education. In J. Cenoz, S. May, & D. Gorter (Eds.), *Language awareness and multilingualism* (pp. 263–280). Cham: Springer.

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. G. Penteado, & O. Skovsmose (Eds.), *Inclusive mathematics education* (pp. 357–375). Cham: Springer.

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.

Greefrath, G., Kaiser, G., Blum, W., & Borromeo Ferri, R. (2013). Mathematisches Modellieren: Eine Einführung in theoretische und didaktische Hintergründe. In R. Borromeo Ferri, G. Greefrath, & G. Kaiser (Eds.), *Mathematisches Modellieren für Schule und Hochschule* (pp. 11–37). Wiesbaden: 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). Hillsdale, NJ: Lawrence Erlbaum Associates.

Haag, N., Heppt, B., Roppelt, A., & Stanat, P. (2015). Linguistic simplification of mathematics items: Effects for language minority students in Germany. *European Journal of Psychology of Education, 30*(2), 145–167.

Haag, N., Heppt, B., Stanat, P., Kuhl, P., & Pant, H. A. (2013). Second language learners' performance in mathematics: Disentangling the effects of academic language features. *Learning and Instruction, 28*, 24–34.
Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold Publishers Ltd.
Hammond, J., & Gibbons, P. (2005). Putting scaffolding to work: The contribution of scaffolding in articulating ESL education. *Prospect, 20*(1), 6–30.
Hasemann, K., & Stern, E. (2002). Die Förderung des mathematischen Verständnisses anhand von Textaufgaben: Ergebnisse einer Interventionsstudie in Klassen des 2. Schuljahres. *Journal für Mathematik-Didaktik, 23*(3-4), 222–242.
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.
Heinze, A., Herwartz-Emden, L., & Reiss, K. (2007). Mathematikkenntnisse und sprachliche Kompetenz bei Kindern mit Migrationshintergrund zu Beginn der Grundschulzeit. *Zeitschrift für Pädagogik, 53*(4), 562–581.
Heppt, B., Stanat, P., Dragon, N., Berendes, K., & Weinert, S. (2014). Bildungssprachliche Anforderungen und Hörverstehen bei Kindern mit deutscher und nicht-deutscher Familiensprache. *Zeitschrift für Pädagogische Psychologie, 28*(3), 139–149.
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.
Jude, N. (2008). *Zur Struktur von Sprachkompetenz.* Wolfgang-Goethe-Universität, Frankfurt.
Kaiser, G. (2017). The teaching and learning of mathematical modeling. In *Compendium for research in mathematics education* (pp. 267–291). Reston, VA: Cai, Jinfa.
Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31.
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.
Kempert, S., Schalk, L., & Saalbach, H. (2019). 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, 66*, 176–195.
Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research, 79*(3), 1168–1201.
Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*(1), 109–129.
Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363–394.
Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K., & Sodian, B. (2015). The development of scientific thinking in elementary school: A comprehensive inventory. *Child Development, 86*(1), 327–336.
Krajewski, K. (2008). Prävention der Rechenschwäche. In W. Schneider & M. Hasselhorn (Eds.), *Handbuch der Pädagogischen Psychologie* (pp. 360–370). Göttingen: Hogrefe.
Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction, 19*(6), 513–526.

Kullberg, A., Kempe, U. R., & Marton, F. (2017). What is made possible to learn when using the variation theory of learning in teaching mathematics? *ZDM Mathematics Education, 49*(4), 559–569.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

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.

Lenhard, A., & Lenhard, W. (2011). Computerbasierte Intelligenzförderung mit den "Denkspielen mit Elfe und Mathis". Vorstellung und Evaluation eines Computerprogramms für Vor- und Grundschüler. *Empirische Sonderpädagogik, 3*(2), 105–120.

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

Lenth, H., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Package 'emmeans': R package version 1.3.1. Retrieved from http://CRAN.R-project.org/package=emmeans

Lewis, A. B. (1989). Training students to represent arithmetic word problems. *Journal of Educational Psychology, 81*(4), 521–531.

Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79*(4), 363–371.

Lewis, C. (1982). *Using the "thinking-aloud" method in cognitive interface design*. Yorktown Heights, NY: IBM.

Liu, R.-D., Wang, J., Star, J. R., Zhen, R., Jiang, R.-H., & Fu, X.-C. (2018). Turning potential flexibility into flexible performance: Moderating effect of self-efficacy and use of flexible cognition. *Frontiers in Psychology, 9*(646), 1–10.

Lubin, A., Rossi, S., Lanoë, C., Vidal, J., Houdé, O., & Borst, G. (2016). Expertise, inhibitory control and arithmetic word problems: A negative priming study in mathematics experts. *Learning and Instruction, 45*, 40–48.

Maier, H., & Schweiger, F. (1999). *Mathematik und Sprache: Zum Verstehen und Verwenden von Fachsprache im Mathematikunterricht*. Wien: ÖBV & HPT.

Markovits, H., & Thompson, V. (2008). Different developmental patterns of simple deductive and probabilistic inferential reasoning. *Memory & Cognition, 36*(6), 1066–1078.

Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Educational Review, 78*(2), 333–368.

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment, 14*, 160–179.

Mayring, P. (2014). Qualitative content analysis: Theoretical foundation, basic procedures and software solution. Retrieved from https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). New York: Macmillan Publishing Company.

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. In L. Puig & A. Gutierrez (Eds.), *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 385–390). Valencia: PME.

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

Meyer, M., & Prediger, S. (2012). Sprachenvielfalt im Mathematikunterricht: Herausforderungen, Chancen und Förderansätze. *Praxis der Mathematik in der Schule, 54*(45), 2–9.

Meyer, M., & Tiedemann, K. (2017). *Sprache im Fach Mathematik*. Berlin: Springer.

Modestou, M., & Gagatsis, A. (2013). A didactical situation for the enhancement of meta-analogical awareness. *The Journal of Mathematical Behavior, 32*(2), 160–172.

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. (2007). Using two languages when learning mathematics. *Educational Studies in Mathematics, 64*(2), 121–144.

Moschkovich, J. (2015). Academic literacy in mathematics for English Learners. *The Journal of Mathematical Behavior, 40*, 43–62.

Mücke, S. (2007). Einfluss personeller Eingangsvoraussetzungen auf Schülerleistungen im Verlauf der Grundschulzeit. In K. Möller, P. Hanke, C. Beinbrech, A. K. Hein, T. Kleickmann, & R. Schages (Eds.), *Qualität von Grundschulunterricht* (pp. 277–280). Wiesbaden: VS.

Muth, K. D. (1984). Solving arithmetic word problems: Role of reading and computational skills. *Journal of Educational Psychology, 76*(2), 205–210.

Muth, K. D. (1992). Extraneous information and extra steps in arithmetic word problems. *Contemporary Educational Psychology, 17*(3), 278–285.

National Governors Association Center for Best Practices, C. o. C. S. S. O. (2010). Common core state standards (Mathematics Standards). Retrieved from http://corestandards.org/

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.

Niederhaus, C., Pöhler, B., & Prediger, S. (2015). Relevante Sprachmittel für mathematische Textaufgaben: Korpuslinguistische Annäherung am Beispiel Prozentrechnung. In E. Tschirner, O. Bärenfänger, & J. Möhring (Eds.), *Deutsch als fremde Bildungssprache: Das Spannungsfeld von Fachwissen, sprachlicher Kompetenz, Diagnostik und Didaktik* (pp. 135–162). Tübingen: Stauffenburg.

Nuerk, H.-C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2015). Extending the mental number line: A review of multi-digit number processing. *Zeitschrift Für Psychologie, 219*(1), 3–22.

OECD. (2013). *PISA 2012 results: Excellence through equity: Giving every student the chance to succeed (Volume II)*. Paris, France: OECD Publishing.

Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review, 22*(2), 115–121.

Páchová, A., & Vondrová, N. (2021). The effect of semantic cues on the difficulty of word problems and the interplay with other complicating variables. *Research in Mathematics Education, 23*(1), 85–102.

Padberg, F., & Büchter, A. (2015). *Einführung Mathematik Primarstufe: Arithmetik*. Berlin: Springer.

Paetsch, J., Felbrich, A., & Stanat, P. (2015). Der Zusammenhang von sprachlichen und mathematischen Kompetenzen bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift für Pädagogische Psychologie, 29*(1), 19–29.

Paetsch, J., Radmann, S., Felbrich, A., Lehmann, R., & Stanat, P. (2016). Sprachkompetenz als Prädiktor mathematischer Kompetenzentwicklung von Kindern deutscher und nicht-deutscher Familiensprache. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 48*(1), 27–41.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193–203.

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). Rotterdam: Sense Publishers.

Paulus, C. (2009). Die "Bücheraufgabe" zur Bestimmung des kulturellen Kapitals bei Grundschülern. Retrieved from http://psydok.psycharchives.de/jspui/handle/20.500.11780/3344

Peng, P., & Lin, X. (2019). The relation between mathematics vocabulary and mathematics performance among fourth graders. *Learning and Individual Differences, 69*, 11–21.

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.

Plath, J., & Leiss, D. (2018). The impact of linguistic complexity on the solution of mathematical modelling tasks. *ZDM Mathematics Education, 50*(1), 159–171.

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.

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.

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.

Prediger, S., & Pöhler, B. (2015). The interplay of micro-and macro-scaffolding: An empirical reconstruction for the case of an intervention on percentages. *ZDM Mathematics Education, 47*(7), 1179–1194.

Prediger, S., & Wessel, L. (2013). Fostering German-language learners' constructions of meanings for fractions: Design and effects of a language-and mathematics-integrated intervention. *Mathematics Education Research Journal, 25*(3), 435–456.

Prediger, S., & Wessel, L. (2018). Brauchen mehrsprachige Jugendliche eine andere fach- und sprachintegrierte Förderung als einsprachige? *Zeitschrift für Erziehungswissenschaft, 21*(2), 361–382.

Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E., & Benholz, C. (2018). Language proficiency and mathematics achievement. *Journal für Mathematik-Didaktik, 39*(2), 77–104.

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.

Quintero, A. H. (1983). Conceptual understanding in solving two-step word problems with a ratio. *Journal for Research in Mathematics Education, 14*(2), 102–112.

Radatz, H. (1983). Untersuchungen zum Lösen eingekleideter Aufgaben. *Journal für Mathematik-Didaktik, 4*(3), 205–217.

Radatz, H., Schipper, W., Ebeling, A., & Dröge, R. (1996). *Handbuch für den Mathematikunterricht*. Hannover: Schroedel.

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

Renkl, A., & Stern, E. (1994). Die Bedeutung von kognitiven Eingangsvoraussetzungen und Lernaufgaben für das Lösen von einfachen und komplexen Textaufgaben. *Zeitschrift für Pädagogische Psychologie, 8*(1), 27–39.

Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist, 44*(2), 162–169.

Resnick, L. B., & Singer, J. A. (1993). Protoquantitative origins of ratio reasoning. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 107–130). Hillsdale, NJ: Lawrence Erlbaum.

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). New York: 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.

Roelcke, T. (2010). *Fachsprachen*. Berlin: Erich Schmidt.

Scheibling-Sève, C., Pasquinelli, E., & Sander, E. (2020). Assessing conceptual knowledge through solving arithmetic word problems. *Educational Studies in Mathematics, 103*, 293–311.

Schindler, V., Moser-Opitz, E., Cadonau-Bieler, M., & Ritterfeld, U. (2019). Überprüfung und Förderung des mathematischen Fachwortschatzes der Grundschulmathematik: Eine empirische Studie. *Journal für Mathematik-Didaktik, 40*(1), 1–35.

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

Schlager, S. (2020). *Zur Erforschung des Zusammenhangs zwischen Sprachkompetenz und Mathematikleistung. Oberflächlichkeit als potenzieller Mediator*. Wiesbaden: Springer.

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly, 23*(2), 139–159.

Schneeberger, M. (2009). *Verstehen und Lösen von mathematischen Textaufgaben im Dialog* (Vol. 529). Münster: Waxmann.

Schneider, W., Küspert, P., & Krajewski, K. (2013). *Die Entwicklung mathematischer Kompetenzen*. Paderborn: Schöningh.

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

Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint). *Journal of Education, 196*(2), 1–38.

Schumacher, R. F., & Fuchs, L. S. (2012). Does understanding relational terminology mediate effects of intervention on compare word problems? *Journal of Experimental Child Psychology, 111*(4), 607–628.

Schütte, M. (2009). *Sprache und Interaktion im Mathematikunterricht der Grundschule: Zur Problematik einer Impliziten Pädagogik für schulisches Lernen im Kontext sprachlich-kultureller Pluralität*. Münster: Waxmann.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge, UK: Cambridge University Press.

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment, 11*(2), 105–126.

Sievert, H., Van den Ham, A.-K., & Heinze, A. (2021). The role of textbook quality in first graders' ability to solve quantitative comparisons: A multilevel analysis. *ZDM Mathematics Education*, 1–15.

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, UK: 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.

Steinbring, H. (1998). Elements of epistemological knowledge for mathematics teachers. *Journal of Mathematics Teacher Education, 1*(2), 157–189.

Stephany, S. (2021). The influence of reading comprehension on solving mathematical word problems: A situation model approach. In A. Fritz, E. Gürsoy, & M. Herzog (Eds.), *Diversity dimensions in mathematics and language learning* (pp. 370–395). Berlin: De Gruyter.

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. (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*. Lengerich: Pabst Science Publishers.
Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word problems. *Cognitive Development, 7*(2), 259–268.
Swan, M. (2003). Making sense of mathematics. In I. Thompson (Ed.), *Enhancing primary mathematics teaching* (pp. 112–124). Philadelphia: Open University Press.
Tarelli, I., Schwippert, K., & Stubbe, T. C. (2012). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, H. Wendt, O. Köller, & C. Selter (Eds.), *TIMSS 2011: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 247–267). Münster: Waxmann.
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.
Thompson, I. (1999). Mental calculation strategies for addition and subtraction. Part 1. *Mathematics in School, 28*(5), 2–4.
Torbeyns, J., De Smedt, B., Stassens, N., Ghesquière, P., & Verschaffel, L. (2009). Solving subtraction problems by means of indirect addition. *Mathematical Thinking and Learning, 11*, 79–91.
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*, 81–117.
Ufer, S., Reiss, K., & Mehringer, V. (2013). Sprachstand, soziale Herkunft und Bilingualität: Effekte auf Facetten mathematischer Kompetenz. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. J. Vollmer (Eds.), *Sprache im Fach* (pp. 185–202). Münster: Waxmann.
Van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorf & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Lawrence Erlbaum Associates.
Van den Ham, A.-K., & Heinze, A. (2018). Does the textbook matter? Longitudinal effects of textbook choice on primary school students' achievement in mathematics. *Studies in Educational Evaluation, 59*, 133–140.
Van Dooren, W., De Bock, D., Vleugels, K., & Verschaffel, L. (2010). Just answering… or thinking? Contrasting pupils' solutions and classifications of missing-value word problems. *Mathematical Thinking and Learning, 12*(1), 20–35.
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.
Verboom, L. (2010). „Ich habe 3 Plättchen mehr als du". *Grundschule Mathematik, 25*, 6–7.
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.
Verschaffel, L., & De Corte, E. (1997). Word problems: A vehicle for promoting authentic mathematical understanding and problem solving in the primary school. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 69–97). Hove: Psychology Press.
Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4*(4), 273–294.
Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology, 84*(1), 85–94.

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, England: Oxford University Press.

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

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education, 52*(1), 1–16.

Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology, 77*(4), 829–848.

Vicente, S., Orrantia, J., & Verschaffel, L. (2008). Influence of situational and mathematical information on situationally difficult word problems. *Studia Psychologica, 50*(4), 337–356.

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.

vom Hofe, R. (1995). *Grundvorstellungen mathematischer Inhalte*. Heidelberg: Spektrum Akademischer Verlag.

Von Damnitz, F. (2020). *Evaluation von Lerngelegenheiten durch Textaufgaben in Schulbüchern der ersten und zweiten Jahrgangstufe: Eine quantitative Bestandsanalyse für den Mathematikunterricht.* (Thesis for the 1. Staatsexamen for primary school education). Ludwig-Maximilians-Universität, München.

Wang, A. Y., Fuchs, L. S., & Fuchs, D. (2016). Cognitive and linguistic predictors of mathematical word problems with and without irrelevant information. *Learning and Individual Differences, 52*, 79–87.

Warner, L. B., Alcock, L. J., Coppolo Jr., J., & 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). Honolulu, Hawaii: PME.

Wartha, S., & Schulz, A. (2011). *Aufbau von Grundvorstellungen (nicht nur) bei besonderen Schwierigkeiten im Rechnen*. Kiel: IPN.

Weber, C. (2016). *Analyse des flexiblen Umgangs mit situationsspezifischen Grundvorstellungen zu Addition und Subtraktion in Sachkontexten unter Berücksichtigung von Familiensprache, Sprachstand und textaufgabenbezogener Mathematikkompetenz: Eine quantitative Erhebung zum mathematischen Situationsverständnis von Drittklässlern.* (Thesis for the 1. Staatsexamen for primary school education). Ludwig-Maximilians-Universität, München.

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

Wessel, J. (2015). *Grundvorstellungen und Vorgehensweisen bei der Subtraktion: Stoffdidaktische Analysen und empirische Befunde von Schülerinnen und Schülern des 1. Schuljahres* (Vol. 23). Wiesbaden: Springer.

Wessel, L., & Erath, K. (2018). Theoretical frameworks for designing and analyzing language-responsive mathematics teaching-learning arrangements. *ZDM Mathematics Education, 50*(6), 1053–1064.

Wilhelm, N. (2016). *Zusammenhänge zwischen Sprachkompetenz und Bearbeitung mathematischer Textaufgaben: Quantitative und qualitative Analysen sprachlicher und konzeptueller Hürden*. Wiesbaden: Springer.

Wilson, P. H., Sztajn, P., Edgington, C., & Myers, M. (2015). Teachers' uses of a learning trajectory in student-centered instructional practices. *Journal of Teacher Education, 66*(3), 227–244.

Xu, L., Liu, R.-D., Star, J. R., Wang, J., Liu, Y., & Zhen, R. (2017). Measures of potential flexibility and practical flexibility in equation solving. *Frontiers in Psychology, 8*(1368), 1–13.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458–477.

Zawojewski, J. (2013). Problem solving versus modeling. In R. Lesh, P. Galbraith, C. Haines, & A. Hurford (Eds.), *Modeling students' mathematical modeling competencies* (pp. 237–243). Dordrecht: Springer.

Zwiers, J., Dieckmann, J., Rutherford-Quach, S., Daro, V., Skarin, R., Weiss, S., & Malamut, J. (2017). Principles for the design of mathematics curricula: Promoting language and content development. *Retrieved from https://ell.stanford.edu/sites/default/files/u6232/ULSCALE_ToA_Principles_MLRs__Final_v2.0_030217.pdf*.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

Gabler, Laura

Name, Vorname

München, 10.08.21 Laura Gabler

Ort, Datum Unterschrift Doktorand*in