LINGUISTIC AND ARITHMETIC FACTORS IN WORD PROBLEMS

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Gabriella Daróczy

aus Miskolc

Ungarn

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	23.02.2021
Stellvertretender Dekan:	Prof. Dr. József Fortágh
1. Berichterstatter:	Prof. Dr. Hans-Christoph Nürk
2. Berichterstatter:	Prof. Dr. Hartmut Leuthold

ABSTRACT

Word problems are among the most difficult problems in mathematical learning. The difficulty of a word problem arises from the task characteristics of the problem itself (e.g., linguistic and arithmetic characteristics), the characteristics of the individual solver (e.g., reading abilities, arithmetic abilities, and intelligence), and the environment (e.g., teaching quality). However, the relation of task characteristics, individual characteristics and environmental factors to the problem-solving process is not yet fully understood. None of the existing models of the word problem-solving process provides a clear answer as to why simple linguistic or numerical modifications that do not change the underlying mathematical structure of the problem affect the solution process. In addition, the role of linguistic and arithmetic task characteristics that determine the difficulty of a word problem have not been systematically investigated so far. Finally, on a theoretical level, existing models do not agree on whether the first reading phase and the calculation phase are separable from other problem-solving processes or if they overlap. Therefore, the aim of this dissertation was to investigate how various task and individual characteristics and environmental factors relate to the problem-solving process of both adults and children and to generate a theoretical process model.

This dissertation distinguishes between (1) unrelated linguistic and arithmetic factors that are conceptually independent of each other, and (2) related factors that are conceptually dependent on each other. To better understand the processes involved in solving word problems, eye-movement data was used to disentangle if the linguistic properties of a problem affect the arithmetic factors, or vice versa. The results of this dissertation show that for both adults and children task characteristics directly influence word problem performance and that linguistic and arithmetic factors interact, indicating that text comprehension and calculation are not necessarily sequential independent processes but partially rely on the same processes during word problem-solving. However, the problem-solving phases seem to overlap more in children than adults, because for adults only related factors interacted, whereas for children both related and unrelated factors interacted. Finally, in addition to task characteristics, individual capability and environmental factors also influenced overall performance. Specifically, children with higher fluid intelligence, reading and mathematical skills, and children from classrooms characterized by cognitive activation and a supportive climate both performed better on word problems. In closing, all factors considering task, individual and environment were integrated into a theoretical process model.

ZUSAMMENFASSUNG

Textaufgaben zählen zu den schwierigsten Herausforderungen des mathematischen Lernens. Die Schwierigkeit einer Textaufgabe ergibt sich durch die Eigenschaften der Aufgabe an sich (z.B. linguistische und arithmetische Eigenschaften), die Eigenschaften der lösenden Person (z.B. Lese-, Rechenfähigkeit und Intelligenz) und die Umwelt (z.B. Unterrichtsqualität). Jedoch ist der Einfluss von Aufgaben-, individuellen Eigenschaften und Umweltfaktoren auf den Problemlöseprozess noch nicht vollständig geklärt. Keines der existierenden Modelle zur Lösung von Textaufgaben liefert eine klare Antwort auf die Frage, warum einfache linguistische oder numerische Modifikationen, die die zugrundeliegende mathematische Struktur des Problems nicht verändern, den Lösungsprozess beeinflussen. Außerdem wurde die Rolle von linguistischen und arithmetischen Aufgabeneigenschaften, die die Schwierigkeit einer Textaufgabe bestimmen, noch nicht systematisch untersucht. Zuletzt widersprechen sich existierende Modelle auf theoretischer Ebene darin, ob die erste Lesephase und die Rechenphase von anderen Problemlöseprozessen getrennt sind oder ob sie sich überschneiden. Daher war das Ziel dieser Dissertation herauszufinden, wie verschiedene Aufgaben-, individuelle Eigenschaften und Umweltfaktoren mit dem Problemlöseprozess bei Kindern und Erwachsenen zusammenhängen, und ein theoretisches Prozessmodell zu entwickeln.

Diese Dissertation unterscheidet zwischen (1) unrelatierten linguistischen und arithmetischen Faktoren, die konzeptionell voneinander unabhängig sind und (2) relatierten Faktoren, die konzeptionell voneinander abhängig sind. Für das bessere Verständnis der Prozesse, die beim Lösen von Textaufgaben involviert sind, wurden Augenbewegungsdaten genutzt, um herauszufinden, ob die linguistischen Eigenschaften einer Aufgabe die arithmetischen Faktoren beeinflussen oder umgekehrt. Die Ergebnisse dieser Dissertation zeigen, dass Aufgabeneigenschaften die Leistung bei Textaufgaben sowohl bei Kindern als auch bei Erwachsenen direkt beeinflussen und dass linguistische und arithmetische Faktoren interagieren, was besagt, dass Textverständnis und Rechnen nicht notwendigerweise sequentielle unabhängige Prozesse sind, sondern teilweise auf den gleichen Prozessen beim Lösen von Textaufgaben beruhen. Allerdings scheinen sich die Phasen des Problemlösens bei Kindern mehr zu überschneiden als bei Erwachsenen, da bei Erwachsenen nur relatierte Faktoren, bei Kindern hingegen relatierte und unrelatierte Faktoren interagierten. Schließlich beeinflussten individuelle Fähigkeiten und Umweltfaktoren zusätzlich zu den Aufgabeneigenschaften die Leistung. Genauer gesagt zeigten Kinder mit höherer fluider Intelligenz, mit besseren Lese- und Rechenfähigkeiten und Kinder aus von kognitiver

Aktivierung und unterstützendem Umfeld geprägten Klassen bessere Leistung bei Textaufgaben. Am Ende wurden alle die Aufgabe, das Individuum und die Umwelt betreffenden Faktoren in ein theoretisches Prozessmodell integriert.

TABLE OF CONTENTS

ABBREVIATIONS

A	addition
ACC	accuracy
ANOVA	analysis of variance
C	consistent form
Ca	carry/borrow
CAOI	combined area of interest
I	inconsistent form
M	mean
nCa	non-carry/non-borrow
N	number of participants
N	nominalized form
ROI	region of interest
RT	reaction time
S	subtraction
SD	standard deviation
SE	standard error
SECO	semantic congruence model
V	verbalized form
WM	working memory
WP	word problem

List of Figures

General Introduction

Study 1

Study 2

Study 3

Study 4

General Discussion

LIST OF TABLES

Study 4

General Discussion

General Introduction

Word problems (WPs) are one of the most well researched topics in mathematics education and have been key components of mathematical education for centuries. WPs usually involve a narrative of some sort which must be mentally translated into an arithmetic problem and subsequently solved (Bassok, Chase, & Martin, 1998). The reason why WPs enjoy such focus is twofold: First, WPs are defined as an essential fundamental to mathematical problem-solving. Second, many people experience severe difficulties in solving WPs (Hegarty, Mayer, & Green, 1992; Lewis & Mayer, 1987; Verschaffel, De Corte, & Pauwels, 1992). For example, according to OECD (2013) one third of students in all participating countries and economies were not able to solve simple one-step WPs involving whole numbers. However, the role of the exact processes (i.e., the process of reading and calculation) and factors (i.e.; how the text is formulated, the background of the solver) associated with problem-solving are often not clear. Surprisingly, it is not just people with low cognitive abilities in reading and calculation who struggle; even expert mathematician can fail to solve some linguistically modified problems (Gros, Sander, & Thibaut, 2019). In addition, many people can complete equivalent calculations when not embedded in text (Verschaffel & De Corte, 1993); therefore, it is not only difficulty that is responsible for poor performance but also the linguistic components of the problem. However, the connection between linguistic and arithmetic factors in particular is not yet fully understood partly because linguistic complexity and arithmetic complexity in WPs are often seen as subsequent (i.e., first, linguistic complexity is experienced through text comprehension then arithmetic complexity through calculation) additive processes which do not interact with each other. Nevertheless, a comprehensive theory of WP solving is still missing (Passolunghi & Pazzaglia, 2004) and different linguistic and arithmetic factors are not differentiated in current models of problem-solving. Different factors might affect the process of WP solving to different extents and may even be processed differently. For instance, a simple change in the text cannot cause the same change in difficulty as changing the whole semantic structure of the WP, where the mathematical structure behind the texts gets more complicated.

Additionally, individual cognitive abilities such as reading skills (e.g, Boonen, de Koning, Jolles, & van der Schoot, 2016), working memory (e.g., Swanson & Fung, 2016), mathematical ability (e.g., Fuchs, Gilbert, et al., 2016) intelligence (e.g., Fung & Swanson, 2017; Xin & Zhang, 2009) play an important role in WP success; nevertheless, which abilities determine performance on different types of WPs is an open question. For instance, linguistic capabilities could be more important for linguistically complex WPs while arithmetic capabilities could be more important for arithmetically complex WPs, and in some problems, both could play a role.

Finally, it is not only the formulation of the text and the individual background of the solver that matters but also the extensive role of the learning environment and this cannot be neglected from problem-solving models. Surprisingly, there are indications that in real life students can solve mathematical problems that they could not in the classroom (Lave, 1992). One explanation is that the teacher might facilitate keyword searching strategies over analysing the text and creating a problem. These simple keyword searching strategies are accused of being behind why many people make errors when solving complex WPs. Altogether it appears that the environment influences how students solve different tasks and the way WPs are taught in the classroom environment may negatively influence student performance. It is unclear how teaching style and teacher-student interaction in the classroom affect performance on different types of WPs.

The dissertation project set out to investigate the underlying cognitive mechanism of WP solving, focusing in particular on the role of different linguistic and arithmetic task characteristics and their relation to problem-solving processes. In this dissertation I investigated a distinction that has not been prominent thus far: the distinction of arithmetic information and linguistic information (i.e., distinguishing factors) which is related or non-related to arithmetic/linguistic; meaning, some linguistic and arithmetic factors are not separable from each other. Namely, unrelated linguistic and arithmetic factors that are conceptually independent of one another, in the sense that they can also be manipulated independently, and related factors which are intrinsically linked. Additionally, besides individual characteristics, this work aimed to examine the relationship between teaching quality and diverse types of WPs. This work is important because the identification of mechanisms underlying problem-solving processes could lead to the development of targeted interventions that may improve educational practice.

THE PROBLEM-SOLVING PROCESS

There have been several models proposed for problem-solving processes: the schema model (e.g., Cummins, Kintsch, Reusser, & Weimer, 1988; Kintsch & Greeno, 1985; Van Dijk & Kintsch, 1983), the situation model (e.g., Thevenot, Devidal, Barrouillet, & Fayol, 2007), the cyclic model (e.g., Bergqvist & Österholm, 2010), and the semantic models (e.g., Bassok, 2001; Gros, Thibaut, & Sander, 2020). The models try to address several key aspects of how people first read, understand

and create a mathematical model – which serves the mathematization (or mathematical representation) of the text – and choose a calculation which in the end they carry out.

- The first aspect is how the mathematical mental representation (here, problem model) is created and what the origin of the mathematical representation is.
- The second aspect involves problem-solving processes. For WPs these processes were named according to Kintsch and Greeno (1985): problem translation (first or initial reading), problem integration, solution planning and solution execution (calculation). During the translation phase (here, initial reading process), "the problem solver constructs an individual mental representation for each sentence of the problem". During the integration phase, the problem solver integrates information across sentences. During the planning phase (here, calculation process), the problem solver develops a plan for solving the problem. The integration and planning phases are usually considered to be where the problem model is created. Finally, during the execution phase, the problem solver carries out the computations called for in the plan. A critical question is how these problem-solving processes interact and/or overlap.
- The third aspect is where errors originate from. Lewis and Mayer (1987) found that students make solving errors because they misunderstand the problem and develop an incorrect representation, not because they conduct a miscalculation. This view is predominant in existing models and most of them see the origins of error as a simple representational source.
- The fourth aspect is a linguistic and arithmetic aspect. Do the models answer why simple changes in linguistic and arithmetic difficulty – where the underlying mathematical model is unaffected – alter performance? Especially, do the models answer why simple arithmetic/numerical changes – i.e., problem size, different types of numbers, carry/non-carry – can change performance?

In the ***schema model*** a mental representation is a propositional domain-specific representation that underlies the problem-solving process (Kintsch & Greeno, 1985). Kintsch and Greeno (1985) adopted the propositional model for WPs, including one significant step: the creation of a problem model. The problem model relies only on schemas stored in long-term memory, the problem-solver's previous knowledge and the text. In this model, the text formulation determines the difficulty of creating an adequate representation (Abedi & Lord, 2001; Cummins et al., 1988). This model sees the phases of problem-solving as clearly separable and serial in which the problem-solver first comprehends the problem situation and performs the required computation: they just fill out the specific numerical values in the schemas. According to the model, errors originate from an incorrect mental representation. The limitation of this model is that it cannot explain why simple rewording (e.g., Davis-Dorsey, Ross, & Morrison, 1991; De Corte, Verschaffel, & De Win, 1985), or changes in numerical values, alters performance (e.g., two- and three-digit numbers, Thevenot & Barrouillet, 2015) because these changes clearly could not affect the schemas.

The ***situation model*** as introduced by Van Dijk and Kintsch (1983) the previously schema-based problem model construction to include the situation. The situation model "corresponds to a level of representation that specifies the agents, actions, and relationships between events in everyday contexts". In the situation model ad hoc transient mental representations are constructed for each problem with the help of working memory (for a summary: Thevenot & Barrouillet, 2015). The prior knowledge of the problem solver and the model is stored in short-term memory. The process of problem-solving is sequential and episodic and problem-solving processes do not overlap. Errors originate from representational errors, not by miscalculations, as the model does not foresee that it is itself influenced by numerical information. The limitations of the situational model are that it does not account for the possibility of the problem-solver creating a new representation and it also cannot explain simple changes in text-only textual changes where the situation model is affected (e.g., situational rewordings).

The ***cyclic model*** from Bergqvist and Österholm (2010) is also based on the situation model – the representation method is the same — but extends it with the possibility of re-representation. According to this model, the first reading creates a mental representation of the text through a process of comprehension but does not have to include the reading of the whole text from the beginning to the end (so the process has overlapping phases). The problem model can include the explicit answer to the question asked in the task or a strategy (e.g., an algorithm or a strategic step „re-read the item") when the problem model is fragmented. This model focuses heavily on how language (especially the first reading process) influences mathematical problem-solving. Errors also originate from false representation; yet, the cyclic model does not address if and how arithmetic processes affect the problem model, so little is known about if this model could answer why simple linguistic and arithmetic changes affect performance.

The ***semantic models*** – i.e., semantic alignment model (Bassok, 2001), semantic congruence model (SECO, Gros et al., 2020) – focus in particular on semantic content, because the situation model does not model the constraints imposed by world semantics on the encoding of arithmetic

WPs In other words, the semantic models extend the method of representation. In the semantic alignment theory of (Bassok, 2001) the semantic content of a problem statement is crucial to a solver. These semantic relations between problem elements thus constrain the solver's representations of the problem, but this model does not include mathematical aspects. The model of Gros et al. (2020) extends semantic alignment theory, provides a distinction between mathematical semantics and word semantics and includes the possibility of the re-representation of the problem model. In this model, word semantics is characterized by the solver's non-mathematical, daily-life knowledge about the elements of the problem statement as well as the relationship between them. On the other hand, mathematical semantics is characterized by the solver's mathematical knowledge applicable to the problem statement. This second model also includes the possibility of the re-representation of the problem model when the initial encoding is incorrect. However, the limitations of both models are that they do not entirely explain when a simple arithmetic change might lead to changes in solution success. Although the model of Gros et al. (2020) provides an explanation for two-step addition, where the problem type was three-step problems instead of two-step problems (e.g., Thevenot & Barrouillet, 2015), but in this case several mathematical semantic representations were available. The models do not account for when the calculation is easy, and no other mathematical representation is available (e.g., in the case of the factor carry / non-carry). Similarly, both models explain only when the linguistic manipulation of the text (e.g., semantic rewording) is affected by the semantics and capitalize on prior knowledge of the solver. Yet, the models do not explain why linguistic manipulations that do not affect the underlying semantic structure (e.g., nominalization) influence solution accuracy (ACC).

In sum, the models can provide explanations when the factors are connected to the semantic, or the underlying, mathematical structure, but fail when linguistic and arithmetic factors do not cause a change in the underlying mathematical structure, or do not influence the connected semantics (i.e., when these factors are unrelated). The models also explain relatively little about the exact relationship between linguistic and arithmetic task characteristics and the process of problem-solving. If problem-solving models are applicable to all different kinds of task characteristics then they should also be able to explain simple cases. It is likely that task characteristics affect problem-solving processes differently.

In addition, the models do not address in particular if and how the processes – i.e., initial reading, creation of the problem model, calculation – overlap. Like the cyclic model, the semantic model also comprises a cyclical component (i.e., when the initial encoding of a problem statement does not lead to a satisfactory algorithm for solving, a recoding may happen to encode a new representation); however, these models hint towards more of an overlapping of the transition and integration phases, not the initial reading and calculation phases. Importantly, these different phases of problem-solving and the separation of the initial reading and calculation phases from the other phases have been postulated but not tested empirically. In addition, it is unclear whether the process of reading is completely separable from the processes of solving. However, whether problem-solving processes are distinct or rely on joint processes may be reflected in the interaction of linguistic and arithmetic task characteristics (Daroczy, Meurers, Heller, Wolska, & Nürk, 2020). Such an interaction could also cause the computational requirements of the problem to interfere with the problem representation (Thevenot & Barrouillet, 2015). Reading comprehension plays a part in almost all of the presented models, especially in those including semantic theories. These models indicate that text formulation affects not only the first reading but also other phases in the solution process. Yet, the models do not address if and how the arithmetic processes affect the problem model. Generally, in all of the models the last process of calculation is seen as distinct because it is suggested that WPs and calculations represent distinct domains of mathematical performance (Fuchs et al., 2008; Swanson & Beebe-Frankenberger, 2004); but, number representation and arithmetic processes might also involve earlier processes. For instance, the study of Munez, Orrantia, and Rosales (2013) suggests that during arithmetic "WP solving, solvers construct a magnitude-based mental representation that goes beyond a conceptual representation in the form of propositions". The only model that considers arithmetic factors is that of Gros et al. (2020). Nevertheless, it explains cases where the arithmetic factor is related to mathematical semantic features but does not provide an answer when the mathematical manipulation does not affect the mathematical semantics (i.e., the calculation is more difficult, as in the case of the operation carry/borrow).

Finally, where errors originate from is also not fully covered by the models as none of them assume the possibility of miscalculation. Some models (i.e., Bergqvist & Österholm, 2010; Gros et al., 2020) include the component of re-representation in the case of an initial inappropriate interpretation. For example, the explanation given by Gros et al. (2020) was that when the calculation was more difficult the schematic representation was affected, and this re-representation was necessary. In the case of arithmetic factors, where schematic representation is unaffected, it is not clear if an initial incorrect representation necessarily exists or if miscalculations are possible. According to the situational model (which is the basis of most of the models), the representation is stored in the working memory. This means that, at the same time, a problem model and the calculation possess the working memory and as the calculation gets more demanding some processes must exist to lessen this pressure on the working memory. There are two possibilities: one is to search for a new representation which enables the calculation to become easier. This is possible according to the Gros et al. (2020) model, that argues for a new re-representation; but, in this case an alternative representation should be available. In other cases where an alternative representation is not available this would mean that the first problem model is not that strong; it is more fragile, but not necessarily wrong. Thus, re-reading the text serves to strengthen (and in a way to memorize) the problem model such that demand on the working memory lessens. If the initial reading and the calculation phases overlap with the other processes the above is possible and would provide a more flexible way of dealing with cognitive resources and explain the cases of unrelated factors.

In sum, the models are not clear on whether the first reading and calculation phases are separable from other problem-solving processes, and do not necessarily consider the joint difficulty of numerical and text factors and their involvement in mental problem-solving.

TASK CHARACTERISTICS

It is necessary to manipulate both linguistic and arithmetic factors independently in order to conclude whether increased difficulty is due to arithmetic complexity, linguistic complexity or the requirement of more cognitive resources. Additionally, a finely graded distinction is important because there are a wide range of different factors in and it is nearly impossible that they influence the solving-process to the same extent. In this dissertation I investigated a distinction that has not been prominent thus far: for instance, both linguistic and arithmetic factors can be divided into two groups: related and unrelated factors. Relatedness means that some linguistic and arithmetic factors are not separable from one another. For example, the keyword which hints towards the solution in a WP is a linguistic factor that can only be defined by the mathematical operation necessary to solve the text problem (e.g., Van der Schoot, Arkema, Horsley, & van Lieshout, 2009). On the other hand, the arithmetic factor operation can also be easily connected to linguistic features in WPs as the operation is defined via the language used. The opposite is true for unrelated linguistic and arithmetic factors, where each remain unaffected by the other. For example, changing the name of the protagonist in a WP (Abedi & Lord, 2001) does not affect the underlying mathematical model. Similarly, whether the numbers in a WP are three-digit or two-digit numbers (Thevenot &

Oakhill, 2006) does not, in principle, change the text. The relationship between related / unrelated factors and problem-solving phases is not clear; however, it is an important distinction because different processing stages should be associated with related, not unrelated, factors. However, none of the problem-solving models above explain how unrelated factors affect the problem-solving process, meaning that they should be extended or modified.

LINGUISTIC FACTORS UNRELATED TO ARITHMETIC FACTORS – NOMINALIZATION

In this dissertation, an unrelated linguistic factor is *nominalization*. *Nominalization* is the process of turning verbs into nouns, and actions into nouns (Francis, 1989), thus increasing the number of noun-phrases in the text (e.g., the verb *"to earn"* with its nominalized form *"the earning"*). WPs using nominalization differ in terms of their grammatical difficulty, but both forms of the problem can be solved with the same arithmetic operation, therefore nominalization is an unrelated factor. It is known that nominalization increases the difficulty of a text (Michael Halliday, Matthiessen, & Matthiessen, 2014; To, Lê, & Lê, 2013) outside the domain of WPs because it has been suggested that uncommon vocabulary in mathematical tasks is negatively related to performance (Abedi & Lord, 2001; Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006). As some simple modifications of the sheer text of a WP without changing the mathematical structure can also lead to lower solution ACC (Cummins et al., 1988; Davis-Dorsey et al., 1991; Staub & Reusser, 1992; Stern & Lehrndorfer, 1992; Vicente, Orrantia, & Verschaffel, 2007) nominalization is a potential influential factor on WP performance.

This idea is supported by Schlager, Kaulvers, and Büchter (2017) who have shown that WPs containing a high density of nominalization significantly affect the solution rate for 10th grade students.

Nevertheless, nominalization does not change the underlying mathematical structure and does not introduce additional arithmetic complexity into the identification of the required calculation. Therefore, this factor should be associated stronger with the initial reading phase.

ARITHMETIC FACTORS UNRELATED TO LINGUISTIC FACTORS – CARRY / BORROW

There is little research regarding basic numerical processing for arithmetic WPs. Although, arithmetic complexity has rarely been systematically considered as an isolated factor, in such WPs evidence from research supports the idea that arithmetic complexity also affects the solving of WPs. Namely, choosing a successful strategy might also depend on the types of numbers in the problem (De Corte, Verschaffel, & Pauwels, 1990), and more difficult calculations might lead to a lower solution ACC (Thevenot & Oakhill, 2006).

One important factor in simple calculation problems, which can be studied in isolation, is the presence or absence of a carry / borrow operation. Children and adults alike take more time and commit more errors when computing the solution to a sum for which adding the units leads to a change in the number of tens (e.g., 14 + 39 = 53) (Deschuyteneer, De Rammelaere, & Fias, 2005; Furst & Hitch, 2000) than when it does not (e.g., 41 + 12 = 63). The borrow operation in subtraction is the inverse operation to the carry operation in addition. The difficulty of two-digit addition and subtraction increases whenever a carry or borrow operation is required (Artemenko, Pixner, Moeller, & Nuerk, 2018). Carry effect – carry effects were more demanding for children when put into a WP (Dresen, Pixner, & Moeller, 2020) and other numerical effects such as the problem size effect (Thevenot & Oakhill, 2006), which suggests that arithmetic problems are harder if their problem size is larger, have been replicated in WPs.

In sum, carrying and borrowing increases the arithmetic complexity of a problem while remaining unrelated to any linguistic property. Therefore, this factor should be more strongly associated with the calculation phase.

LINGUISTIC FACTORS RELATED TO ARITHMETIC FACTORS – LEXICAL CONSISTENCY

One major linguistic factor referring to the relationship between text and arithmetic in a problem is lexical inconsistency (de Koning, Boonen, & van der Schoot, 2017; Pape, 2003; Van der Schoot et al., 2009; Verschaffel et al., 1992), which was introduced by Lewis and Mayer (1987). Lexical consistency concerns specific keywords in the text, so-called cue words, which signal or hint towards particular arithmetic operations like addition or subtraction (Hinsley, Hayes, & Simon, 1977). Therefore, this factor can be seen as related to the correct arithmetic operation. A WP is considered "lexically consistent" if the semantics of the cue word signals an operation that is congruent with the operation required for the correct solution. For example: *"A man saves money on some purchases. He had 82 euros. He earns 15 euros. How much money does the man have now?"* Here the cue word earns is associated with getting more; that is, it infers the operation can be associated, in this case, with addition, and the solution indeed requires addition: $82 + 15$. In lexically inconsistent problems, the operation hinted at by the lexical meaning of the cue words is inconsistent with the operation needed for the correct solution. For example: *"A man saves money on some purchases. He spent 74 euros. He has 23 euros now. How much money did the man have?"* Here the cue word spent is commonly associated with having less than before; thus, it is associated with subtraction. However, in this example, computing the solution requires addition: 74 + 23 = 97 – examples are from Daroczy et al. (2020). Verschaffel et al. (1992) found a consistency effect for children but not for adults when they presented one-step arithmetic compare problems. The consistency effect has been found to affect the transition and integration phases of the problem-solving process but not the initial reading phase (Verschaffel et al. 1992).

In sum, lexical consistency increases the complexity of a problem and is related to arithmetic; therefore, this factor should be more strongly associated with phases where the problem model is built.

ARITHMETIC FACTORS RELATED TO LINGUISTIC FACTORS – OPERATION

Most WP research addresses addition and / or subtraction as the arithmetic operation (e.g., Carpenter, Hieben, & Moser, 1981; De Corte & Verschaffel, 1987; Fennema et al., 1996; García, Jiménez, & Hess, 2006; Klein, Nuerk, Wood, Knops, & Willmes, 2009; K. Moeller, E. Klein, & H. C. Nuerk, 2011b). In regards to the effect of operation, it has been shown that subtraction is generally more difficult (Artemenko, Moeller, Huber, & Klein, 2015) and elicits greater reaction times (RTs) than addition does (Orrantia, Rodríguez, Múñez, & Vicente, 2012). Additionally, some types of subtraction tasks can be solved by various strategies, including indirect addition or direct subtraction (De Corte & Verschaffel, 1987; Torbeyns, De Smedt, Ghesquière, & Verschaffel, 2009). There is reliable evidence outside the WP literature that the operation influences arithmetic performance of both children (e.g., Dresen et al., 2020) and adults (e.g., Orrantia et al., 2012).

Both operation and lexical consistency should be related to one another; therefore, they should be associated more with the phases of problem-solving where the problem model is created (i.e., integration and transition phases). This assumption is consistent with previous literature suggesting both operation and lexical consistency affect the second stage of problem-solving.

INTERACTION OF LINGUISTIC AND ARITHMETIC FACTORS

Understanding the possible interactions between linguistic and arithmetic factors enables the dissociation of the origin of problem difficulty. For instance, the resolution of linguistic and arithmetic difficulties may rely on the same processing stages and resources (Sternberg, 1969), like working memory. Since working memory affects all components of the complexity of a WP, the difficulties triggered may not be simply additive, but interactive. If both the linguistic and arithmetic elements of a problem are complex this might lead to particularly slow processing as working memory has limited resources. However, it is impossible to use RT to investigate if linguistic factors influence arithmetic factors or if arithmetic factors influence linguistic factors, whereas eye-movements – eye-movement research proved to be useful in the detection of the mechanisms underlying difficulties associated with the linguistic and arithmetic properties of a WP located in various parts of the text (De Corte et al., 1990; Hegarty et al., 1992; Van der Schoot et al., 2009; Verschaffel et al., 1992) –, for instance, can indicate greater attention to the text in more difficult mathematical conditions and enables the identification of the parts of the text that require greater attention. If there is an interaction or common processing stage, linguistic / arithmetic difficulty should lead to a different eye-movement pattern over those regions of the problem which correspond to the other part of the problem (i.e., linguistic elements for greater mathematical difficulty and numerical elements for greater linguistic difficulty). This means that not only numerical areas, but also the textual areas, should be affected by a change in the eye-movement pattern if linguistic difficulty changes. In sum, the interaction between linguistic and arithmetic factors should also be present in eye-movement behaviour.

In addition, the hypothesis of the dissertation is that related and unrelated factors may not have the same effect on different solving processes. For example, the factor operation (subtraction / addition) is more heavily dependent on mental representation because the language that operation is expressed in depends on the semantics and semantic knowledge is linked to the conceptual knowledge needed to solve problems, such as knowledge about increases or decreases in quantity (Rabinowitz & Wooley, 1995). The interaction between the two related factors – i.e., lexical consistency and operation – should therefore be especially pronounced. I expect this interaction to be over-additive – i.e., when the relationship between operation and consistency is such that both factors are difficult, due to limited resources. Because complex linguistic and arithmetic task characteristics make a task more difficult they can increase domain-general attributes such as cognitive load (Daroczy et al., 2020). Additionally, I hypothesize longer RTs than expected based on the main effects only. However, I would expect interactions not only between related factors but also between unrelated and other factors, but less consistently than in the case of related factors at. The exact reason why unrelated factors interact less in all processing stages is that they should affect less the stage where the problem model is built due to their independence from semantic features. Finally, these factors are also more distinct from the other processing stages.

Finally, exploring the presence or absence of an interaction between related and unrelated factors could also help to understand the problem-solving process. This would also serve to investigate if problem-solving phases are sequential, because the existing models predict different interactions. There should be no interaction between related linguistic factors / arithmetic factors and unrelated linguistic factors / arithmetic factors according to the propositional theory, as it sees the initial reading phase, the calculation phase and the building of the problem model phase as distinct problem-solving phases. According to Sternberg (1969), interactions are then not possible because there is no common stage of processing for the cognitive processes underlying the manipulated factors. In contrast to the propositional model, the models that consider problem-solving as sequential and cyclic would suggest an interaction between related and unrelated factors because the manipulated factors might operate partially in a common stage of processing (Daroczy et al., 2020). In this case, interaction between the unrelated carry / borrow factor and any other factors would suggest that the calculation process interacts with other problem-solving phases. Similarly, an interaction between an unrelated nominalization factor and any other factors would suggest that the reading comprehension phrase interacts with other problem-solving phases (Daroczy et al., 2020). Contrary to this, factors that are related to other arithmetic / linguistic factors should affect not only the first reading and the calculation phases, but also the phase where the problem model is built. This is supported by the fact that an interaction between operation and lexical consistency is often found (e.g.; Van der Schoot et al., 2009; Verschaffel et al., 1992). The word problem-solving models, regardless of allowing for overlapping processes or not, would suggest different eye-movement patterns. For example, according to the propositional theory from Kintsch (1980) greater attention should not be paid to the numbers in the case of nominalization / consistency and solvers should not look back more frequently at the numbers either. Only the operation should affect the numbers in the text because it is involved in the calculation process. On the other hand, according to the models which argue for the non-sequential nature of problem-solving, linguistic factors should also affect numerical areas. This indicates that calculation processing might also influence another stage of problem-solving and be part of the mental representation. Similarly, the text without the numbers, should be unaffected by the calculation phases, according to the propositional theory. This means that the factor operation should play less of a role than any linguistic factors. However, in a case where operation would influence eye-movement behaviour over the text the process of problem-solving text comprehension is confounded by the calculation process when the difficulty of the operation increases.

In sum, both linguistic and arithmetic complexity contribute to the difficulty of a WP, but likely to different extents. In the theoretical models we should differentiate between related and unrelated factors as this could also help to gain deeper insight into the process of problem-solving.

INDIVIDUAL CHARACTERISTICS

Above and beyond linguistic and arithmetic complexity, the relationship between text and an arithmetic problem, cognitive factors and individual factors also influences the ease and likelihood of correctly solving WPs. There is vast literature on the background of problem solvers, in terms of reading abilities, mathematical abilities, intelligence and working memory. Nevertheless, individual characteristics are not integrated into models of WP solving. There is one model applicable to children or adults; however, individual characteristics might influence WP solving in two different ways. The first one is a direct effect, the second one is via an interaction through mediating factors.

A direct effect of individual characteristics on WP solving would mean that such characteristics would generally affect performance independently of problem type, which is indeed the case for many individual characteristics. For instance, several studies show a direct influence on WP solution, which means that lower language ability compared to better language ability leads to worse performance on WPs (Andersson, 2007; Chow & Ekholm, 2019; Fuchs, Geary, Fuchs, Compton, & Hamlett, 2016; Jordan, Levine, & Huttenlocher, 1995; M. O. Martin & Mullis, 2013). In addition, Thevenot and Barrouillet (2015) and Kail and Hall (1999) showed that both reading competence and mathematical competence have a significant impact on the solution rate for WPs; indeed, mathematical competence is a necessary foundation for WPs (Fuchs, Fuchs, Compton, Hamlett, & Wang, 2015). Fluid intelligence has been shown to predict WP performance (Mori & Okamoto, 2017; Swanson & Beebe-Frankenberger, 2004) and it has been suggested that WP performance is generally related to working memory (e.g., Passolunghi & Siegel, 2001; Swanson, Cooney, & Brock, 1993). It has been shown that for other domain specific and domain general factors working memory plays a role not only for children (e.g., the verbal and spatial components of working memory Soltanlou, Pixner, & Nuerk, 2015) but also for adults. Additionally, motivation not only positively influences mathematical performance in general (Gaspard et al., 2016) but also WP performance (Gasco & Villarroel, 2014) and mathematical performance. This is supported by Möller, Pohlmann, Köller, and Marsh (2009), who found a correlation between self-concept and mathematics performance in schools. Last, but not least the OECD (2013) has shown that more socioeconomically advantaged students compared to students with lower socioeconomic status score higher (the equivalent of nearly one year of schooling) in mathematics. This finding seems to be applicable to the domain of WP solving because there is a positive relationship documented between socioeconomic status and WP solution ACC (Jordan, Kaplan, Nabors Oláh, & Locuniak, 2006). These results should be easily replicated in this dissertation.

The direct influence of individual characteristics on WP solving does not fully explain why some individuals can solve one type of WP and others cannot. For example, Nortvedt, Gustafsson, and Lehre (2016), reported a strong positive correlation between numeracy and reading ability. Another example is that studies have shown that the correlation between working memory and problem-solving ACC is lower when differences in reading comprehension are corrected for (e.g., Fuchs et al., 2006). Individual characteristics might influence WP solving performance over mediator variables. One important mediating factor could be cognitive load. Cognitive load is perceived mental effort and refers to the amount of working memory used; heavy cognitive load can have negative effects on performance (Sweller, 1988). In addition, the interaction between linguistic and arithmetic factors probably increases working memory demands, thus subjects with better working memory compared to lower working memory should perform better. Children and adults might experience a different amount of cognitive load and cognitive load for an individual with excellent linguistic or arithmetic abilities may be lower for the same problem than for an individual with poor linguistic or arithmetic abilities. Indeed, linguistically simple items reduce the cognitive load associated with language demand in WPs, but this seems to depend on the language abilities of the problem-solver (Pennock-Roman & Rivera, 2011). Therefore, participants with better reading skills should be less affected by linguistic complexities than participants with poor reading skills. At the same time, arithmetic abilities are not often referred to or measured (Boonen, van der Schoot, van Wesel, de Vries, & Jolles, 2013) but mathematical skills are assumed to be closely linked to working memory. A task which is demanding in terms of arithmetic processing is also demanding in terms of the whole solution process. Indeed, the experience of cognitive load is not the same for everyone and this could lead to different solution strategies. Individual characteristics could influence individual WP performance over solution strategies. This is supported by the finding that, like in the domain of WPs, working memory capacity influences the choice of solution strategies for WPs (Thevenot & Oakhill, 2006). In addition, differences in motivation and self-concept seem to influence the strategies used to solve WPs (Gasco & Villarroel, 2014). Even simple reading strategies can differ (Rayner, 1998). For example, De Corte and Verschaffel (1986) suggest a relationship between the ability of children to solve WPs and their initial reading pattern. The authors found that most rereading was done by highly capable children. However, it is important to note that in WPs of lower semantic difficulty less successful individuals were also able to apply successful strategies (Van der Schoot et al., 2009). Therefore, it is not only the task characteristics, but also the connection between the task characteristics and individual abilities that is important, so I hypothesise that individual differences together with task characteristics should affect WP performance.

In sum, I hypothesise that individual differences together with task characteristics should affect WP performance both directly and over mediator variables like cognitive load and individual strategies. Cognitive load for an individual with excellent linguistic or arithmetic abilities may be lower for the same problem than for an individual with poor linguistic or arithmetic abilities. Finally, solution strategies may facilitate WP solving if specific solution strategies are applied to a WP because the problem type allows this and because the individual knows the strategy.

ENVIRONMENT

There is one crucial component which has not yet been considered in existing problem-solving models. Failures of many students to engage in problem-solving behaviour can not only be attributable to poor comprehension skills or cognitive deficits, instead they might originate, for example, from pedagogical factors (e.g., Ulu, 2017). For example, some studies (e.g., De Corte & Verschaffel, 1987; Lean, Clements, & Del Campo, 1990) indicate that the way WPs are taught in the classroom environment might have an influence on performance and chosen strategies. Teaching method and quality of instruction are especially critical for proper learning and play an important role in academic performance (Thatcher, Fletcher, & Decker, 2008). This should apply not only to mathematical achievement in general, but also to WP success (Nortvedt et al., 2016). Therefore, the environment cannot be neglected when developing WP solving models. In a study by Nortvedt et al. (2016) a positive correlation between instructional quality and mathematical achievement was observed. More specifically in the domain of WPs, solving competence depends heavily on the quantity and quality of the instruction (for review see Verschaffel, Schukajlow, Star, & Van Dooren, 2020). How teachers organize the classroom affects mathematical performance (Fuchs et al., 2015; Peterson, Fennema, Carpenter, & Loef, 1989; Suviste, Kiuru, Palu, & Kikas, 2016) and false assumptions about WPs (e.g., that every problem can be solved with a single operation; Cummins et al., 1988) are sometimes believed and instructed by teachers (Fuentes, 1998; Nortvedt et al., 2016). This indicates a positive relationship between teaching quality and performance on WPs. However, it is unclear how teaching quality affects the WP performance because the effect of teaching quality on WP solving has not been directly investigated.

Teaching quality – introduced by Klieme, Pauli, and Reusser (2009) is the conceptualization of the interaction between students and their teachers in the classroom. According to Klieme et al. (2009) teaching quality comprises three factors: supportive climate, cognitive activation and structured classroom management. Supportive climate covers "specific aspects of the teacher student relationship such as positive and constructive teacher feedback, a positive approach to student errors and misconceptions, and caring teacher behavior" (Fauth, Decristan, Rieser, Klieme, & Büttner, 2014). Cognitive activation includes challenging tasks, and the exploration of concepts, ideas and prior knowledge (Lipowsky et al., 2009). Classroom management focuses on classroom rules and procedures, coping with disruptions and smooth transitions, and affects student achievement (Fauth et al., 2014; Seidel & Shavelson, 2007). The hypothesis of this dissertation is that each of the three factors of teaching quality should affect WP solution ACC. The first subscale, cognitive activation, should play a role because teachers might facilitate cognitive activation (Verschaffel, De Corte, & Borghart, 1997). Teachers even might propagate the keyword strategy (e.g., Boote & Boote, 2018; Jitendra et al., 2015), for which students do not necessarily need to create a proper problem model (Hegarty, Mayer, & Monk, 1995) of the task and instead search only for keywords and numbers followed by carrying out one operation. Of course, such strategies might work for simple problems, but not for more demanding ones. Nevertheless, keyword strategies are more error prone when WPs get more complicated, for example, those that are lexically inconsistent (Hegarty et al., 1992; Hegarty et al., 1995). Therefore, cognitive activation should influence WP success because it encourages problem understanding, prior to the use of the keyword strategy, a merely stepwise, and simple rule-based, direct translation strategy. The hypothesis of this dissertation is that in classrooms that encourage cognitive activation teachers use tasks that challenge students to create problem models. In such classrooms greater cognitive activation should decrease the inconsistency effect. This also means that environmental factors are connected to task characteristics. Classroom organization, support offered to students and explanations of how to solve WPs have a great impact on learning (Titsworth, McKenna, Mazer, & Quinlan, 2013). Student achievement is affected by classroom management because classrooms that experience few disruptions (i.e., classrooms with good classroom management), providing an environment where students are better able to concentrate on their tasks, are expected to be associated with WP solution ACC. Classrooms with a positive and supportive climate generally enhance student performance (Hugerat, 2016). Therefore, the hypothesis is that this subscale of teaching quality should have a positive impact on WP performance because when students working with text problems are better supported and instructed by their teachers they are likely to be better equipped to deal with difficulties.

These findings indicate the way WPs are taught in the classroom plays an essential role in solution success; however, it is unclear how teaching quality affects WP performance. In sum, higher quality teaching should lead to greater solution ACC and better mathematical performance. Additionally, there should be a positive relationship between teaching quality scales and difficult problems, and therefore a relationship between task characteristics and the environment. Last but not least, environmental factors should be considered in WP solving models.

Objectives of the Studies

The current dissertation project aimed to investigate the underlying cognitive mechanism of WP solving, focusing in particular on the role of different linguistic and arithmetic task characteristics and their relation to problem-solving processes including individual characteristics and environmental factors (i.e., teaching quality). It is important to understand how various task and individual characteristics and environmental factors relate to the process of problem-solving in order to culminate this information into a theoretical process model. To better understand the process of problem-solving, this dissertation distinguishes between related and unrelated linguistic and arithmetic factors. This distinction is important because none of the existing WP solving models give a clear answer as to why very simple linguistic or arithmetic modifications affect solution processes. The existing models fail in particular to explain cases in which linguistic and arithmetic factors do not cause a change in the underlying mathematical structure, or do not influence the semantics (i.e., when both factors are unrelated to each other). Investigating the interactions between linguistic and arithmetic factors could provide further understanding of problem-solving processes as existing models are also unclear on whether the first reading phase and the calculation phase are separable from other problem-solving processes and how the processes overlap. There should be no interaction between related linguistic factors / arithmetic factors or unrelated linguistic factors / arithmetic factors according to the propositional theory but there should be interactions between these factors according to the cyclic model. Finally, the source of error is also not fully covered in existing models as none of them assume the possibility of miscalculation.

Therefore, the first focus of this dissertation project is on the interaction of linguistic and arithmetic task characteristics and the nonassociation between related and unrelated factors. In all experimental studies I manipulated linguistic complexity to be independent and orthogonal to the arithmetic complexity of the arithmetic problem underlying the WP in such a way that there was one factor related to arithmetic / linguistic complexity and one unrelated factor (e.g., Daroczy et al., 2020). For example, the linguistic factor lexical consistency is related to arithmetic complexity, while the linguistic factor nominalization is not; analogously, the arithmetic factor operation is related to linguistic complexity, while the arithmetic factor carry / borrowing is not (see Figure 1). The dissertation explored how the interactions between related and unrelated linguistic and arithmetic factors could support one of the existing problem-solving models.

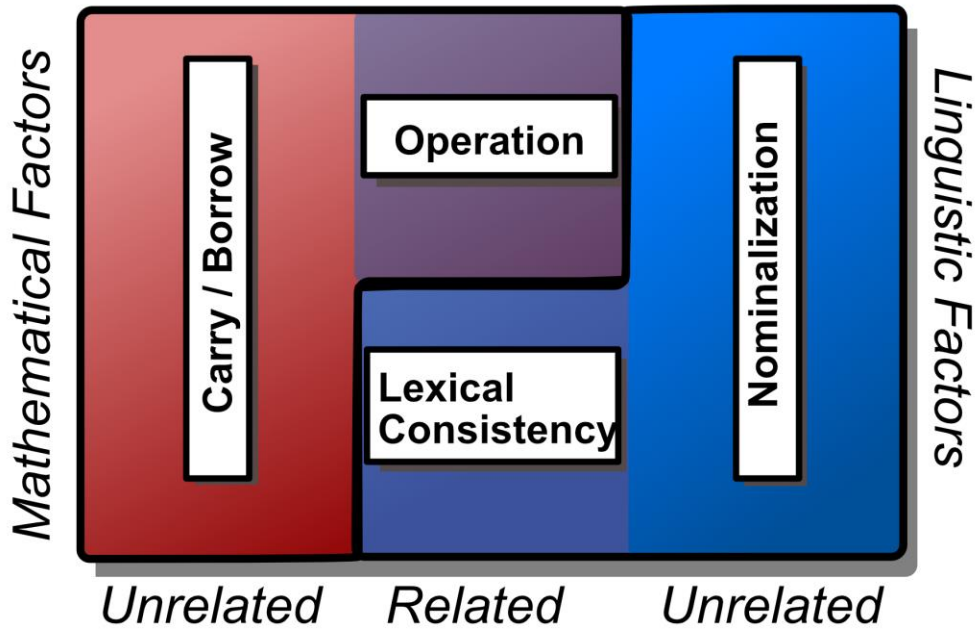

Figure 1 Related and Unrelated Linguistics and Arithmetic Factors.

Furthermore, how individuals with different abilities relate to the process of problem-solving was evaluated. This served as a conceptual replication of previous studies and as controls. What is new and adds to the existing literature is that this dissertation looked in detail at the specific interactions between task characteristics and individual abilities and not only at the main effect. By considering individual characteristics together with environmental factors it is possible to determine why some individuals have more difficulties with one or the other type of problem taught in schools. This is important because different students may have difficulties with different types of WPs.

As stated above, this dissertation also aimed to explore how the environment influences WP solving processes. This is important because existing problem-solving models do not consider environmental factors, however there is much evidence that they influence the problem-solving processes. For example, little is known about if there is a positive relationship between teaching quality scales and linguistically and arithmetically demanding items.

Behavioural and eye-tracking methods were used to understand the cognitive processes underlying different task demands such as linguistic and arithmetic complexity and the contribution from individual factors. Therefore, a series of studies were planned to investigate the questions on the process of problem-solving derived from above:

Theoretical Process Model (Study 1). In a first step, a review about WP solving was addressed, as a comprehensive theory about problem-solving is missing and the existing models usually do not distinguish between linguistic and arithmetic factors. Thereby, this study aimed to explore various linguistic and arithmetic factors. Here I have distinguished three components of WP difficulty: (i) the linguistic complexity of the text of a problem itself, (ii) the arithmetic complexity of the arithmetic problem, and (iii) the relationship between the linguistic and arithmetic complexity of the problem. This study identifies the basics of linguistic and arithmetic concepts, as well as provides a theoretical process model of WP solving which integrates all the components: task characteristics, individual characteristics and environmental factors. In the next steps, (Study 2-3) empirical studies aimed to validate the theoretical process model.

Adult Study (Study 2). This study aimed to provide a better understanding of the processes involved in solving WPs and to clarify if the problem-solving phases overlap/interact. The interactions between related and unrelated linguistic and arithmetic factors were explored experimentally with 25 adult participants. The arithmetic factor was operation (addition vs. subtraction, carry / borrow vs. non-carry / non-borrow); the linguistic factors were lexical consistency (lexically consistent form vs. lexically inconsistent form) and nominalization (verbalized form vs. nominalized form). For linguistic / arithmetic factors which are related to other arithmetic / linguistic factors, such interactions were expected to be particularly pronounced and more consistent (i.e., in the case of lexical consistency and operation). Beyond the main effects I expected an interaction of linguistic and arithmetic complexity. Additionally, I measured mathematical and reading abilities (by means of a speed-reading test and speed-calculation tests in addition and subtraction) as well as working memory (by letter span and Corsi block-tapping test).

Child Study (Study 3). The objectives of this study were addressed with different samples because WP processing differs between children and adults. The results from of adult study are not necessarily transferable to children, because children are still undergoing their education and therefore might use different problem-solving strategies and have different skills, for example, the cognitive load of a given problem might be higher for children than adults. Additionally, children might process problems differently. This study also addressed existing WP solving models, strategies and overlapping phases. I was also interested in if the behavioral results could be reflected in eye-movement patterns. Therefore, to evaluate whether linguistic and arithmetic factors interact I measured the WP solving ability of 34 5^{th}-6^{th} grade children (aged 10-13). The arithmetic factor operation (addition vs. subtraction); the linguistic factors lexical consistency (lexically consistent form vs. lexically inconsistent form) and nominalization (verbalized form vs. nominalized form) were included in the WP stimuli. The factor carry / borrow was kept constant. Specific capabilities (arithmetic, reading, text comprehension and working memory) were assessed. Besides the interaction of related and unrelated linguistic factors, eye-movement behaviour (where attention is focused, which elements of the WPs are re-read and processed more deeply) was investigated. In doing so, the key question is how linguistic and arithmetic difficulty affect the processing of the whole problem as well as of specific parts (e.g., numbers, text) of WPs. Generally, more complex items receive more attention and re-readings. Additionally, I investigated a hypothesis not previously tested: that increasing difficulty of mathematical information shifts eye-movements towards numerical elements and that increasing linguistic difficulty shifts eye-movements towards textual elements of a WP. Moreover, an interaction of linguistic and arithmetic factors is expected: namely, for linguistically complex WPs an increase in attention paid to numerical elements is expected and for arithmetically complex WPs an increase in attention paid to textual elements is expected. Lastly, I investigated how task characteristics influence eye-movement behavior and how individual abilities are related to different reading strategies as it is possible that children are more diverse in their use of reading strategies than adults, as children are just learning to deal with WPs.

Classroom Study (Study 4). The focus of this study was not on which children are the most successful at solving WPs, but which children solve some kinds of WPs better and how this relates to environmental factors. Besides the interaction of related and unrelated linguistic factors, I extended Study 3 to investigate the relationship between teaching quality and various types of WPs of high linguistic and arithmetic difficulty. Therefore, I examined WP solving performance in a sample of 386 5^{th}-6^{th} (aged 10-13) grade students. Specific capabilities (reading, text comprehension, working memory and intelligence), social aspects (socio-economic status), arithmetical capabilities and attitudes (motivation and self-concept) were also assessed. I expected a positive relationship between teaching quality scales and linguistically and arithmetically demanding items (cognitive activation in particular should increase ACC on lexically inconsistent items). To measure teaching quality, I used student ratings, which have been shown to be useful measures of teaching quality in primary school (Fauth et al., 2014; Sandilos, Rimm-Kaufman, & Cohen, 2017).

In sum, the main goal of this dissertation was to broaden our understanding of the key linguistic and arithmetic factors contributing to the difficulties of solving WPs. The central aim of this dissertation was to investigate the theoretical question of how WP solving works by means of the interaction between linguistic and arithmetic factors. Only with such differentiation on an item level (as regards linguistic and arithmetic complexity and their interrelation), on an individual level (as regards linguistic and arithmetic skills and general cognitive abilities) and an environmental level (as regards classroom and teaching attributes) will it be possible to understand why individuals have difficulties with some types of WPs. This integration of all three aspects is especially important for curriculum development and to improve educational practice.

STUDIES

The following chapters are written as separately readable manuscripts. This results in overlapping contents to the chapters introduction, discussion and between the empirical chapters:

STUDY 1: 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*. doi:10.3389/fpsyg.2015.00348

STUDY 2: Daroczy, G., Meurers, W. D., Heller, J., Wolska, M., & Nuerk, H.-C. (2020). The interaction of linguistic and arithmetic factors affects adult performance on arithmetic word problems. *Cognitive Processing*. doi:10.1007/s10339-019-00948-5

STUDY 3: Daroczy, G., Meurers, W. D., Artemenko, C., Wolska, M., & Nuerk, H.-C. (ready for submission). Influence of task characteristics on eye-movement patterns related to numerical and textual information in arithmetic word problems. doi:10.31234/osf.io/dz9he

STUDY 4: Daroczy, G., Fauth, B., Meurers, W. D., Cipora, K., & Nuerk, H.-C. (ready for submission). The relation of environmental factors to the task difficulty in word problems. doi:10.31234/osf.io/dz9he

STUDY 1: WORD PROBLEMS: A REVIEW OF LINGUISTIC AND NUMERICAL FACTORS CONTRIBUTING TO THEIR DIFFICULTY

CONTRIBUTIONS OF CO-AUTHORS AND OTHER PERSONS

Authors

Gabriella Daroczy[1,2], Magdalena Wolska[2], Detmar Meurers[2,3]& Hans-Christoph Nürk[1,2,4]

[1]Department of Psychology, Eberhard Karls University Tuebingen, Tuebingen, Germany

[2]LEAD Graduate School & Research Network, Eberhard Karls University Tuebingen, Tuebingen, Germany

[3]Department of Linguistics, Eberhard Karls University Tuebingen, Tuebingen, Germany

[4] Knowledge Media Research Center, Tübingen, Germany

Author	Author position	Scientific idea (%)	Data generation (%)	Analysis and Interpretation (%)	Paper writing (%)
Daroczy, G.	1	50	100	60	80
Wolska, M.	2	10	0	10	5
Meurers, W. D.	3	20	0	15	7
Nuerk, H.-C.	4	20	0	15	8

Proof-Reading

Lora Minkova

I/We certify that the above-stated is correct.

28.09.2020 Hans-Christoph Nürk

Date, Signature of at least of one of the supervisors

ABSTRACT

Word problems belong to the most difficult and complex problem types that pupils encounter during their elementary-level mathematical development. In the classroom setting, they are often viewed as merely arithmetic tasks; however, recent research shows that a number of linguistic verbal components not directly related to arithmetic contribute greatly to their difficulty. In this review, we will distinguish three components of word problem difficulty: (i) the linguistic complexity of the problem text itself, (ii) the numerical complexity of the arithmetic problem, and (iii) the relation between the linguistic and numerical complexity of a problem. We will discuss the impact of each of these factors on word problem difficulty and motivate the need for a high degree of control in stimuli design for experiments that manipulate word problem difficulty for a given age group.

Keywords

word problems, linguistics complexity, numerical complexity, text properties, difficulty

INTRODUCTION

Word problems (WPs) are part of the school curriculum and are taught at all levels of education. In WPs, relevant information is presented in the form of a short narrative rather than in mathematical notation (Verschaffel, Greer, & De Corte, 2000). Sometimes WPs specifically encode a quantitative relation between objects (Boonen et al., 2013). Many children from kindergarten through adulthood have severe difficulties in solving WPs (Hegarty et al., 1992; Lewis & Mayer, 1987; Riley, Greeno, & Heller, 1983; Verschaffel et al., 1992). Both linguistic and numerical complexity contributes to the difficulty in solving WPs. However, researchers have so far often focused on the one or the other aspect, depending on which field they come from. Even within the respective fields, linguistics and numerical cognition, some aspects have been studied extensively, while others have been (strangely) neglected. For instance, we will see that semantics and discourse structures have been frequently studied in the context of WP complexity, but systematic syntactic manipulations are scarce. As regards numerical cognition, number properties like parity and magnitude as well as the type of mathematical reasoning have often been studied, but the type and the form of operations (e.g., carry-over effects) have not been investigated thoroughly in WPs, although they play an important role in current numerical cognition research (Moeller, Klein, Fischer, Nuerk, & Willmes, 2011; Nuerk, Moeller, Klein, Willmes, & Fischer, 2011; Nuerk, Moeller, & Willmes, 2015).

In this review, as researchers from the field of linguistics and the field of numerical cognition we have collaborated to provide a systematic overview of linguistic and numerical aspects relevant to solving WPs as well as their interaction. To capture a broad range of relevant facets in the review, we extended our view of the relevant literature with systematic keyword searches in journals including the following terms: WPs, story problems in combination with situational model, performance, consistency hypothesis, language processing, relational terminology, semantic influence, rewording, semantic cues, number size and type, working memory, text comprehension, computational errors, operations, position of unknown.

INDIVIDUAL DIFFERENCES AND SOCIAL FACTORS

Social consequences of difficulties in WPs must be considered (Fite, 2002). For example, in the PISA studies—often measured with WPs—, mathematics literacy is a commonly used notion (Stacey, 2015). “It is defined as an individual’s capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that" (OECD, 2010). Unsuccessful WP solvers can experience negative social health, life outcome (Schley & Fujita, 2014). Even beyond social consequences, numerous studies focused on individual differences and group differences, such as students with and without learning disabilities (Kingsdorf & Krawec, 2014), and children with and without developmental disabilities (Neef, Nelles, Iwata, & Page, 2003). Hegarty et al. (1995) distinguished domain-specific strategies that successful and unsuccessful problem solvers develop with practice and how these strategies account for individual differences in performance. Different students – e.g., individuals with calculation difficulty, or WP difficulty (Powell & Fuchs, 2014) – may struggle with different types of WPs.

The *role of social factors*, schooling, teachers, curricula, peers, motivation and how the way of teaching influences solution strategies is not a focus of this review. Nevertheless, mentioning them is important given that they play a role in the solution success. Factors of relevance include the way of responding (e.g., De Corte, 1988), the scoring criteria, the presence of illustrations next to the text (e.g., Berends and Van Lieshout (2009), or solution models used by the teachers. School WPs also support stereotypical thinking: WPs do not resemble problems in real-world situations (Foong & Koay, 1997) and there is a strong tendency both among students and teachers to exclude real-world knowledge from their WP solution (Verschaffel et al., 1997). This can be due to the observation that the problem-solving process is also influenced by social cognitive and epistemic behavior settings (Reusser, 1988). Linguistic and pedagogical factors also affect children's understanding of arithmetic WPs (Lean et al., 1990). Students' beliefs about what doing and knowing mathematics means are rather different from the ideals (Jiménez & Verschaffel, 2014) and shaped by "socio-mathematical norms". Indeed, differences in motivation. The role of social factors, schooling, teachers, curricula, peers, motivation and how the way of teaching influences solution strategies is not a focus of this review. Nevertheless, mentioning them is important given that they play a role in the solution success. Factors of relevance include the way of responding (e.g., De Corte, 1988), the scoring criteria, the presence of illustrations next to the text (e.g., Berends & Van Lieshout, 2009), or solution models used by the teachers. School WPs also support stereotypical thinking: WPs do not resemble problems in real-world situations (Foong & Koay, 1997) and there is a strong tendency both among students and teachers to exclude real-world knowledge from their WP solution (Verschaffel et al., 1997). This can be due to the observation that the problem-solving process is also influenced by social cognitive and epistemic behavior settings (Reusser, 1988). Linguistic and pedagogical factors also affect children's understanding of arithmetic WPs (Lean et al., 1990). Students' beliefs about what doing and knowing mathematics means are rather different from the ideals (Jiménez & Verschaffel, 2014) and shaped by "socio-mathematical norms". Indeed, differences in motivation seem to influence the strategies used to solve WPs (Gasco, Villarroel, & Zuazagoitia, 2014).

SUBCATEGORIES OF WORD PROBLEMS AND SOLUTION STRATEGIES

Several different types of WPs—e.g., in the underlying mathematical structure, solvability— are often presented intermixed in one study without acknowledging the problem type. This is problematic. Different types of WPs are presented for various students' groups, in different schools or different age groups. For example, Swanson, Lussier, and Orosco (2013) investigated the role of strategy instruction and cognitive abilities on WP-solving ACC. The mathematical WPs they used were: addition, subtraction and multiplication without any further description of the problem type. However, the available literature has already shown that different categories of WPs may lead to different solution strategies and different error types. For instance, different semantic problem types result in different errors (Vicente et al., 2007) and have a different difficulty level (LeBlanc & Weber-Russell, 1996). Obviously, different scientific studies reporting results for different student or age groups cannot be easily compared to one another when they use different WP types; it cannot be determined whether differences should be attributed to group or study manipulation or differences in the used stimulus material. In the following, we outline the major distinctions discussed in the literature. Besides the difficulty level, WPs have been categorized with regard to various other attributes. Based on standard algebra text books, Mayer (1981) categorized WPs according to their frequency. Riley et al. (1983) created four groups based on the *semantic structure* of additive arithmetic WPs (change, compare, combine, equalize) and eighteen further subcategories. For instance, the change problem –where there is a start, a change and a result state –can be subdivided into three subcategories depending on which state is the unknown.

The *mathematical content* of WPs can also serve as a basis for categorization. Algebra WPs typically require translation into a mathematical formula, whereas arithmetic WPs are solvable with simple arithmetic or even mental calculation. In contrast to arithmetic WPs, algebraic reasoning WPs share the same numerals and signs (Powell & Fuchs, 2014) and the manipulation of those numbers and signals differs based on the question or expected outcome (Kieran, 1990). However, the distinction is not that straightforward as in some cases both methods can be applied. For instance, in a study by Van Dooren, Verschaffel, and Onghena (2002), future secondary school teachers preferred the use of algebra even when an arithmetical solution seemed more evident, and some future primary school teachers rather applied arithmetical methods. Computer-aided environments have been introduced for algebraic WPs (Reusser, 1993); to support learning on “getting the formalism” and the “equation” (Nathan, Kintsch, & Young, 1992) and to allow students to generate, manipulate and understand abstract formal expressions for WPs. However, solution approaches are not easily dissociable between arithmetic and algebraic problems. If a WP is intended to be solved with an equation, in some cases a simple arithmetic approach is enough (Gasco et al., 2014). Under some circumstances, it is even easier to solve WPs via alternative arithmetic strategies than by deriving algebraic equations. US children perform better on a story problem if it is in a money context and the numbers involve multiples of 25 (Koedinger & Nathan, 2004). While the distinction between algebra and arithmetic WPs is important for investigation and evaluation, in this review we concentrate mainly on arithmetic WPs.

Standardized phrases and the *idea that every problem is solvable* are other important attributes of many, but not all WPs. Textbooks generally suggest implicitly that every WP is solvable and that every numerical information is relevant (Pape, 2003). They usually provide standardized phrases and keywords that are highly correlated with correct solutions (Hinsley et al., 1977). There are so-called non-standard WPs (Jiménez & Verschaffel, 2014) which can be non-solvable WPs or if they are solvable some have multiple solutions and may contain irrelevant data. In the recent literature, non-standard WPs are getting more and more attention (Csíkos, Kelemen, & Verschaffel, 2011) Children give a high level of incorrect answers to non-standard WPS because these seem to contradict their mathematics-related beliefs learned in the classroom. Reusser (1988) presented 97 first and second graders the following sentence: "There are 26 sheep and 10 goats on a ship. How old is the captain?" and 76 students “solved” the problem using the numbers in the task. The rationale behind such studies is that always-solvable textbook problems with standardized phrases and including only relevant numerical information are hardly ecologically valid. Real-life WPs are not standardized, contain irrelevant information, and a solution may not always exist.

The above subcategories, which essentially characterize specific sets of WP properties, have a direct impact on human performance in WP. For space limitations, we cannot discuss the impact of all subcategories in detail, but we illustrate their impact on performance and strategies with two examples: (i) Different subcategories can result in different errors, and involve different representations and processes. For example, a familiar misconception is that multiplication (Vergnaud, 2009) always makes the result larger (which is not true for $n < 1$), that division makes the results smaller, and that division involves always division of the larger number by the smaller.

(ii) Addition problems are strongly influenced (De Corte & Verschaffel, 1987) by the semantic structure (change, compare, combine). Carpenter et al. (1981) reported that the dominant factor in determining the children's solution strategy was the structure of the problems. For example, the subcategories Change 2 and Change 3. With Change we are referring to the classification of Riley et al. (1983) which requires the child to find the difference between the two numbers given in the problem; however, the strategies the children used to solve these problems were quite different. Following examples are adapted from Riley et al. (1983).: Change 2: “Joe had 8 marbles. Then he gave 5 marbles to Tom. How many marbles does Joe have now?” Change 3: “Joe had 3 marbles. Then Tom gave him some more marbles. Now Joe has 8 marbles. How many marbles did Tom give him?”. Almost all the children used a subtraction strategy (e.g.: counting up) to solve Change 2. For Change 3 almost all the children used an addition strategy (e.g.: counting down). Haghverdi, Semnani, and Seifi (2012) suggest that the highest number of errors result among others from a lack of semantic knowledge. In sum, the subcategories introduced in this section influence both performance and the choice of solution strategies.

The solution strategies people develop and their mediating factors have systematically been in the focus of WP research and addressed the following questions: How do children and adults solve WPs? Why do they make different errors and at which level of the solution process they do so? Which kind of semantic representation do they create of the WP? Which skills are necessary for the solution process? The first theories on WP solution processes (Kintsch & Greeno, 1985) have drawn on the text comprehension theories of Van Dijk and Kintsch (1983) and Mayer (1984). When solving problems, the solver first integrates the textual information into an appropriate situation model or a mental representation of the situation being described in the problem, which then forms the basis for a solution strategy. This approach was further applied by (Jiménez & Verschaffel, 2014; Kingsdorf & Krawec, 2014; Thevenot & Oakhill, 2005). An important foundation of those approaches is that solving WPs is not a simple translation of problem sentences into equations (Paige, 1966). Often both WPs and the corresponding numerical problems are done without language translation (Schley & Fujita, 2014). Several researchers have focused on abstraction as a reductive process involved in the translation process in the WPs. Nathan et al. (1992) argue that WPs solving is an exercise in text processing required for understanding the problem (Cummins et al., 1988). WP-solving is highly dependent upon language comprehension skills, however reading and mathematics skills are often viewed as separate factors. Successfully solving WPs has been argued to require at least three distinct processes (Nesher & Teubal, 1975):

(i) Understanding and constructing the relation between text and arithmetic task, (ii) linguistic understanding of the WP itself (iii) solving the arithmetic tasks. Typically, only the latter process is assumed to be shared with common arithmetic tasks. Many students can successfully solve common arithmetic tasks and they show good text comprehension skills. Yet they fail to solve WPs correctly. This suggests that other factors also play a major role for the solution. Besides domain-general capabilities like IQ, the role of domain specific knowledge and processes were investigated to get a complete account of problem-solving, basic cognitive abilities; visual, reading skills, mathematical skills and metacognitive abilities involved in the solution process. For example Boonen, van Wesel, Jolles, and van der Schoot (2014) and Oostermeijer, Boonen, and Jolles (2014) explored the role of spatial ability and reading comprehension in WP-solving, since good WP solvers do not select numbers and relational keywords but create a visual representation (Boonen et al., 2013). The role of solution strategies has been studied extensively for children.

LINGUISTICS COMPLEXITY & LINGUISTICS STUDIES

In linguistics, the notion of complexity is discussed under a range of perspectives, with particularly fruitful definitions grounded in research on language evolution (Nichols, 1990) and language acquisition (Bulté & Housen, 2012). Following the latter, it is useful to delineate linguistic complexity from propositional complexity (the amount of meaning to be expressed) and discourse-interactional complexity (the interaction of participants in discourse). This makes it possible to zoom in on linguistic complexity as the degree to which a text at hand is elaborate and varied (R. Ellis, 2003 p. 340). Linguistic complexity can be analyzed with respect to all aspects of the linguistic system: from the words and their lexical and morphological aspects, via the way these words can be combined in syntax to form sentences, to the text structure and overall discourse. Languages differ with respect to where in the linguistic system complexification is supported. For example, English makes use of word order to encode grammatical functions, whereas agglutinative languages such as Hungarian or Turkish make use of a rich morphological inventory for this and other uses. The implication of linguistic encoding differences is two-fold: First, the difficulty of WPs is language-specific, thus linguistic manipulation leading to increased WP complexity in one language may not have an effect in another, more complex language. Second, the performance of language learners on WPs presented in a foreign language may be affected by the differences between the learner's mother tongue and the language of the problem presentation. In the following two sections, we briefly summarize the main findings on aspects of linguistic complexity that affect performance.

STRUCTURAL FACTORS

Studies on the relation between linguistic structure and student performance on WPs have considered complexity at the micro-level of word and sentence forms as well as at the macro-level of the discourse structure of the WP passage. Early approaches addressed structural complexity in terms of basic quantitative properties of the WP text, such as the number of letters, words, sentences, mean word and sentence length, or the proportion of complex (long) words (Lepik, 1990; Nesher, 1976; Searle, Lorton, & Suppes, 1974). More linguistically-motivated variables have been investigated in the context of comprehension difficulties in WPs for language learners, for the most part learners of English. At the vocabulary level, comprehension difficulties which result in problem-solving difficulties for English language learners may stem from the presence of unfamiliar (low-frequency) words, polysemous words, idiomatic or culturally-specific lexical references. At the sentence structure level, factors that have been shown to play a role include noun phrase length, the number of prepositional phrases and participial modifiers, the presence of passive voice and complex clause structure such as relative, subordinate, complement, adverbial, or conditional clauses (Abedi & Lord, 2001; Abedi, Lord, & Plummer, 1997; Martiniello, 2008; Shaftel et al., 2006; Spanos, Rhodes, & Dale, 1988; Thevenot et al., 2007).

At the discourse structure level, specifically in terms of discourse ordering, the correspondence between the order in which numerical data is presented in the WP and the order in which it can be used to solve it has been shown to be a major predictive variable. Order-consistent problems result in better performance (Searle et al., 1974). Better performance has also been observed for simpler question wording or placing the question before the text results (Cummins et al., 1988).

SEMANTIC FACTORS

A single factor that is straightforwardly related to WP difficulty and that has been widely investigated is the presence or absence of explicit verbal cues whose semantics hint at the expected operation and thus directly lead toward the solution. Verbal cues include words and phrases of different categories: conjunctions („and“for addition), adverbs („left“, „more than“, „less than“ for subtraction), or determiners („each“ for multiplication). Eye tracking studies have shown that subjects tend to focus on linguistic verbal cues and perform translation directly to the mathematical operation (Hegarty et al., 1992; Van der Schoot et al., 2009).

Because verbal cues so often lead to default mathematical interpretation (Nesher, 1976), even small differences in phrasing in cue words can cause significant changes in performance (LeBlanc & Weber-Russell, 1996). This is especially relevant for young children (Lean et al., 1990), who in the course of development connect words such as „join“, „add“, „get“, „find“, or „take away“ with concepts such as putting together, separating, giving away, or losing. A problem can thus be reworded by adding verbal clues which make the semantic relations more salient so that the underlying mathematical relation is more explicit. For example, the WP “There are 5 marbles. Two of them belong to Mary. How many belong to John?” can be reworded as “There are 5 marbles. Two of them belong to Mary. The rest belongs to John. How many belong to John?” (from Cummins, 1991). This kind of conceptual rewording has been shown to be useful to improve children’s performance on WPs (Vicente et al., 2007). Thus changes in wording can influence representation (De Corte et al., 1985).

Semantic or object relations between the objects described in the problem also relate to difficulty. Division problems usually involve functionally related objects (e.g., tulips–vases) and rarely categorically related objects (e.g., tulips–daisies) (S. A. Martin & Bassok, 2005). By contrast, addition for the most part involves categorically related objects. The correlation between object relations and mathematical operations has been argued to reflect a structural correspondence between semantic and mathematical relations (Bassok et al., 1998). For this reason, the semantic structure properties of a WP have been emphasized as a more important factor contributing to difficulty than the syntactic structure (Kaur & Yeap, 2001; S. A. Martin & Bassok, 2005). Interestingly, an effect related to information load has been observed; the presence of content irrelevant to the core solution, i.e., the presence of numerical or linguistic distractors, results in higher error rates (Muth, 1992). De Corte and Verschaffel (1987) found that the semantic structure of WPs influences children’s choice of mathematical solution strategy. In terms of the broader task context, the required or expected way of responding to the WP has a big influence, especially for the domain of multiplication and division with rational numbers as argued in De Corte (1988); for example, whether students are expected to answer the problem numerically or if they only have to indicate the required operation, or whether they respond in an open way or with multiple choice.

NUMERICAL COMPLEXITY & NUMERICAL STUDIES

Arithmetic WPs have to be usually transformed mentally into an arithmetic problem and usually require an arithmetic solution (S. A. Martin & Bassok, 2005). This means transforming word and numbers into the appropriate operation (Neef et al., 2003). Since the arithmetic problem has to be solved in the end, numerical representations and arithmetic processes will also play an important role in the solution process. In numerical cognition, different models and representations have been proposed (e.g., Dehaene & Cohen, 1995; Nuerk et al., 2012). However, the problem here is that the literature on WP often seems (with some exceptions) to be largely in a parallel research universe to the literature on numerical cognition and arithmetic processes, so that standard models of numerical cognition are hard to apply on the existing literature. What is more, WP research on numerical/arithmetic factors is also affected by the scoring criteria; in some studies on WP-solving, computational errors are neglected, because in many studies researchers consider a solution as correct as long as the solver has chosen the correct mathematical model (Verschaffel & De Corte, 1990). This is not the case in behavioral numerical cognition research, where the correct result is usually essential and RTs, accuracies, error types and solution types are analyzed based on the arithmetic problem and result.

Numerical complexity can influence WP performance via at least 3 routes (see Figure 1):

1. Direct route: WPs with more complex arithmetic structure are more difficult independent of linguistic complexity.
2. Cognitive load: More complex arithmetic problems involve a higher cognitive load. For instance, carry problems are supposed to require more working memory resources. If the linguistic properties are also complex and the built-up of a mental model also requires more working memory resources, high linguistic and arithmetic complexities could lead to over additive difficulties which could neither be explained by main effects of linguistic or numerical difficulty.
3. Solution strategies: Multi-digit numbers are harder to process than single-digit numbers (Nuerk et al., 2011, for a review), and arithmetic complexity usually increases with numbers of digits. Thevenot and Oakhill (2005) compared the influence of processing 3-digit numbers and 2-digit numbers on WP solution strategies. They showed that processing numerically more complex 3-digit numbers facilitated alternative strategies by the participants. The authors suggested that higher

work load and working memory led to this facilitation. For our review and the model in the revised manuscript, the important point is that they resort to less effortful strategies. Similar results were observed by Brissiaud and Sander (2010) who manipulated the size and order of the numbers and operation resulting thus in two different solution strategies: (i) Situation-strategy WPs that are easy to solve with informal strategies e.g.: double-counting, derived number fact or trial-error strategy. (ii) Mental Arithmetic-strategy WPs are "easy to solve with mental calculation, but only when the relevant arithmetic knowledge is used". Number magnitude and order determined which strategy was used most likely. In this review, we suggest that resorting to alternative easier strategies is not restricted to number magnitude, but could be used with any numerical variable that allows simpler solutions. For instance, if a number bisection task were used in a text problem, we would also suggest that participants resort to easier strategies (e.g., checking the parities of the outer number), when the bisection problem gets more complex (e.g., larger interval, decade crossing etc.).

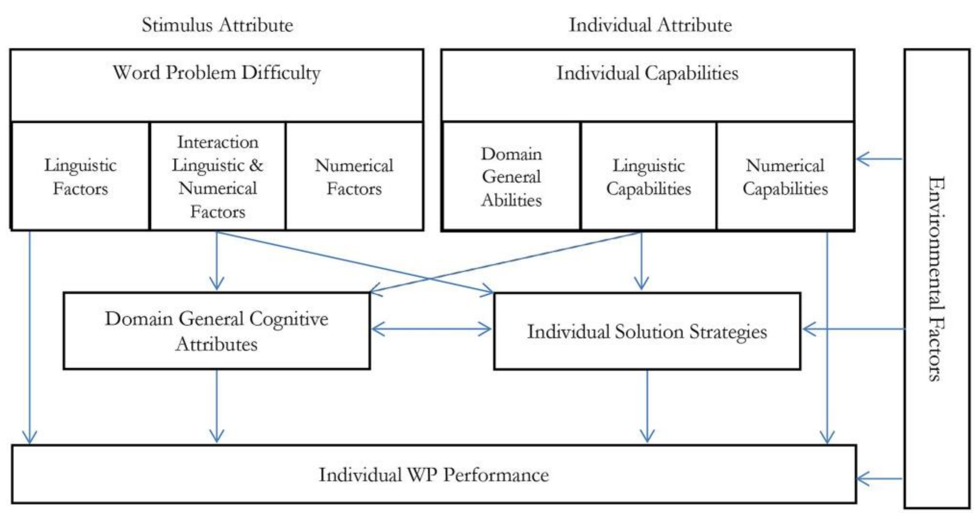

Figure 1 - 1 Theoretical Process Model.

This figure describes a possible theoretical process model of world problem-solving based on this article and dissociating numerical and linguistic factors: Three general aspects are distinguished for predicting individual WP performance. Stimulus Attributes (WP difficulty), individual attributes (capabilities), and environmental factors (e.g., teaching). WP difficulty comprises linguistic factors (such as linguistic complexity of the WP text, Section 2 of this article), numerical/arithmetic factors (such as numerical difficulty of the numerical problem, Section 3), and their interaction (such as the relation between text and arithmetic problem, Section 4). Individual Capabilities can refer to linguistic and numerical capabilities and domain- general abilities such as individual working memory capacity. Stimulus attributes and individual attributes influence individual WP performance both directly and over two mediator variables. One mediator variable refers to domain-general attributes, such as cognitive load. Complex linguistic and numerical stimulus attributes can increase cognitive load and the impact of increased complexity may be over additive, especially when the joint linguistic and numerical complexity exceeds the cognitive load of an individual. On the other hand, those domain- general attributes are influenced by individual capability. Cognitive load for an individual with high linguistic or numerical abilities may be lower for the same problem than for an individual with low linguistic or numerical abilities. The second mediator variable refers to specific solution strategies. If specific solution strategies can be applied to a particular WP problem, because the problem type allows this and because the individual knows the strategy, solution strategies can facilitate WP solving. Finally, environmental factors (e.g.: teaching, scoring system… etc.) influence individual capabilities, solution strategies, and also directly individual WP Performance.

Nevertheless, some distinctions of numerical processes can be made in our review of the WP literature and are therefore proposed as an initial step in this review. Note that in our view this is not the end of the integration of numerical cognitive research and WP research, but rather just a beginning. For an overview of the investigation of specific numerical processes in current WP research, we suggest categorizing them into five categories:

1. the property of numbers (parity, single digit/multi digit, problem size, ties, type of number, role of the number, number magnitude),
2. required operation (type, number)
3. mathematical solution strategies (larger number, place, automatic fact retrieval, position of the unknown),
4. relevance of the information.
5. Other numerical processes and representations.

NUMBER PROPERTIES

While some studies have shown an effect of numerical complexity, from a numerical cognition view it is surprising that actually the arithmetic complexity has rarely been systematically considered as an isolated factor in WPs although it is frequently examined in other arithmetic problems or simply the description of numbers is missing e.g.: Verschaffel and De Corte (1990).

For instance, parity attributes are rarely considered in WPs although in children it influences task performance and strategy choice in arithmetic tasks. For instance, in the number bisection task (Is the middle number Y the exact mean of X and Z in X_Y_Z?), parity influences performance. Trials with unequal parities of X and Z are easier to solve than trials with equal parities (Nuerk, Geppert, van Herten, & Willmes, 2002). We suggested that this is due to a change in strategy. In trials with unequal parity (e.g., 21_25_28), it is impossible that the middle number is the mean, because the mean of numbers with unequal parity is not an integer number (and only integers were used in the experiment). Therefore, participants may change their strategy after they discovered unequal parities and may not compute further to find out whether the middle number is really the mean. A later fMRI study by Wood et al. (2008) corroborated this assumption. In the easier unequal parity ("impossible") condition, we observed more activation in the right ventrolateral prefrontal cortex, which is activated in cognitive set changes or when participants generate alternative solutions for a task. Thus, parity can influence performance and solution strategies in arithmetic. This seems not only the case in the bisection task, which is to our knowledge rarely used in WP research, but also in standard operations like addition and subtraction. A review by Hines (2013) suggests that parity influences the difficulty of addition and subtraction, but not multiplication, and tasks containing odd numbers are more difficult than with even ones. Such parity effects have received little attention in WP research so far. Furthermore, it seems that most WPs, especially for children, contain single-digit numbers; e.g., each answer was in the range of 1-9 e.g. in Lean et al. (1990), or Fuchs et al. (2014), only few use multi-digit numbers (Haghverdi et al., 2012). In Nesher (1976) the range of numbers is smaller than 100, contained division two digit numbers into one digit number).

Explanations why the studies have chosen specific numbers, e.g., mentioning problem size are rare. Verschaffel and De Corte (1990) and Orrantia, Rodríguez, and Vicente (2010) controlled for the number of sentences; the size of the numbers given in the problems. In the study of Van der Schoot et al. (2009) the final answers were between 14 and 40, included no fraction, no negative number, no numerical value twice and none of the possible answers resulted in another. However, different types of numbers were presented in WPs in some more studies: (i) fraction (Raduan, 2010), (ii) whole number, (iii) decimal number (Haghverdi et al., 2012); and their effect has been rarely investigated. Koedinger and Nathan (2004) found an effect for decimal numbers: "However we also observed a smaller situation facilitation effect whereby story performance was better than word equation performance under certain conditions: namely dealing with decimal numbers".

The mixed use of single- and multi-digit numbers is problematic because in the last 15 years, numerous numerical cognition studies have shown that single-digit number processing cannot easily be generalized to multi-digit number processing e.g. Nuerk, Weger, and Willmes (2001); for reviews see Göbel, Moeller, Pixner, Kaufmann, and Nuerk (2014) and Nuerk, Weger, and Willmes (2005). Nuerk and colleagues have identified 17 numerical effects linked to different numerical representation, which are specific for multi-digit number processing and which cannot be explained by single-digit number representations. Also, even the same effects are different for single- and multi-digit numbers. For instance, Ashkenazi, Rubinsten, and Henik (2009) have shown that the distance effect for two-digit numbers differentiates between dyscalculic and typically developing children. The sometimes seemingly arbitrary mix of single-digit and multi-digit number use in WP research is therefore not reasonable in our view given the state of numerical cognition research and the major differences between processing those different number types. The role of a number within an operation also influences WP complexity (De Corte, 1988). For example, in the case of addition the role means addend minuend or by multiplication: multiplicand, multiplier. One important finding from recent research on multiplication WPs is that children's performances are strongly affected by the nature of the multiplier whether e.g.: it is an integer, decimal larger than 1 or a decimal smaller than. On the other hand, the size of the multiplicand has little or no effect on problem difficulty. De Corte (1988) stated that "two multiplication problems with the same mathematical, semantic and surface structure but different in terms of the nature of the given numbers can elicit very distinct levels of problems difficulty". Indeed, this corresponds to recent findings that relatedness and consistency heavily influence the ease with which a multiplication problem can be solved cf. for relatedness (Domahs, Delazer, & Nuerk, 2006; Domahs et al., 2007) and for consistency Verguts and Fias (2005).

Despite the major role of number properties in numerical cognition, number property has not been investigated extensively in the WPs (Fuchs et al., 2009). Nevertheless, numbers seem to play a major role. For instance, De Corte and Verschaffel (1986) observed that in their eye tracking study there is a relatively strong focus on the numbers in the problem. Twenty-five percent of the total solution time was spent in the two small number areas. However, major number properties of numerical cognitions research such as number magnitude are rarely systematically considered in

WP research. In our view, more dialogue between fields, – numerical cognition and WP research – seems necessary.

REQUIRED OPERATION

Carrying out operations are necessary steps in solving arithmetic WPs. Operations have been used extensively in WPs. Most errors seem to originate from people's failure to understand the language of WPs, i.e., the linguistic embedding of the calculation problem (Schumacher & Fuchs, 2012), and arithmetic computation errors themselves (Kingsdorf & Krawec, 2014; Raduan, 2010). Some errors may result from correct calculation performed on incorrect problem representation (Lewis & Mayer, 1987) and different operations may lead to different solution strategies. The most usual operation used in WP experiments is addition and subtraction (Carpenter & Moser, 1984; De Corte, 1988; Schumacher & Fuchs, 2012). Even the classification of Riley et al. (1983) was made for elementary addition and subtraction. Research in the 80s and 90s concentrated on how children learn to do one step addition and subtraction problems involving small whole numbers; see the review from Vicente et al. (2007). Later, the focus was more on the multiplication WPs or mixed WPs – e.g., Swanson and Beebe-Frankenberger (2004). Greer (1992) presented a framework categorization of multiplication and division WPs on the basis of the types of quantities involved (positive integers, fraction and decimals) as models of situation. The semantic problem structure also influences the solution strategies for addition and subtraction.

Choosing the correct operation depends strongly on the type of the given numbers in the problem (De Corte et al., 1990). As already shortly outlined in above subsection on problem types, there is a huge body of research on what makes addition, subtraction or multiplication problems difficult. Carry operations (e.g., 28 + 47; the decade value 1 from the unit sum 15 has to be carried over to the decade sum) have long been known to make multi-digit addition more difficult in children and adults; see (Nuerk et al., 2015) for a review. However, solution strategies differ between children and adults – eye movement data suggest that in a choice reaction task elementary school child always compute and search for the correct results, while adults seem to also decide based on the rejection of the incorrect result. What is more, even within the carry operations at least three different cognitive processes can be identified for adults: Unit sum calculation, carry detection, and carry execution (K. Moeller, E. Klein, & H.-C. Nuerk, 2011a). Inability to execute one of these processes may lead to worse performance in carry problems in particular. What is more, carry addition problems seem to require larger working memory resources (Ashcraft, 1992; Furst & Hitch, 2000). If cognitive load/working memory demand is high, because both the linguistic and the numerical complexity of the WP are large, this may lead to overadditive problems in the domain-general processing stages involved in WP-solving — see Figure 1, for an elaboration. For multiplication, we know that relatedness, ties, whether a problem stems from the 0,1, 2, 5, or 10 row (Jost, Khader, Burke, Bien, & Rösler, 2009), or consistency influence the difficulty of a multiplication problem (Domahs et al., 2006; Domahs et al., 2007). Although such factors have been extensively studied in numerical cognition research, they are – to the best of our knowledge – rarely considered in WP research. Since we know that these factors make the arithmetic computation, which is part of the WP solution, much more difficult, this lack of consideration is again problematic in our view.

MATHEMATICAL SOLUTION STRATEGIES

Mathematical solution strategy variations have been studied extensively, and can be a function of e.g.: wording and semantic categories. Mathematical solution strategy variations have been studied extensively, and can be a function of linguistic factors like wording, semantic categories, propositions. However, how individuals come up with mathematical solution strategies can be also influenced by numerical/arithmetic factors like number magnitude (Thevenot & Oakhill, 2005). Such variables, which are independent of other factors, make WPs harder and/or influence numerical representations have rarely been studied. The position/place of the unknown variable has an effect on representation (García et al., 2006). Even studies about working memory also investigated the position of the unknown variable (Swanson & Beebe-Frankenberger, 2004). The strategy of counting on from larger is easier if the bigger number is represented first (Wilkins, Baroody, & Tiilikainen, 2001). Even for adults: 4+2=6, and 2+4 =6 which are mathematically equivalent, may psychologically imply different meanings (Kaput, 1979). The sequence of the numbers, e.g., whether a problem starts with the smaller or with the larger number (Verschaffel & De Corte, 1990) the position of the numbers and particular words (Schumacher & Fuchs, 2012) influence children's solution of elementary addition and subtraction problems. For example, in change problems children look typically for a specific number to begin with, depending on task features, like the first mentioned number (Lean et al., 1990; Wilkins et al., 2001), the type of problem (start or change set), and the size of the numbers (Verschaffel & De Corte, 1990).

Arithmetic fact retrieval is well researched ubiquitous strategy in numerical cognition but less so in the domain of WPs. Orrantia et al. (2010) found that arithmetic fact retrieval is not limited to simple addition, but also possible in other tasks, such as single-digit arithmetic WPs. Fuchs et al. (2009) investigated so called "Number combination". This means simple arithmetic problems that can be solved via counting or decomposition strategies or committed to long term memory for automatic retrieval. Here, arithmetic fact retrieval had to be differentiated from other strategies on three levels: operational, items difficulty and individual differences. These numerical/arithmetic factors influence solution strategies in arithmetic and WPs as well. Decomposition and counting require more working memory and therefore leave less resources for the built-up and maintenance of a text situations model. However, both individual and stimulus differences should also be considered. For instance, Grabner et al. (2009) showed in an fMRI study that not only problem size, but individual strategy choice contributed to fact retrieval processes when solving multiplications.

<table>
<tr><th>Linguistic Factors</th><th>Mathematical Factors</th><th>General Factors</th></tr>
<tr><td rowspan="2">Structure
Structural complexity of basic quantitative properties
(e.g., Number of letters, word and sentence length, Proportion of complex words)
(Lepik, 1990; Nesher, 1976; Searle et al., 1974)
Vocabulary level
(e.g., polysemous words, prepositional phrases, passive voice, clause structure)
(Abedi & Lord, 2001; Abedi et al., 1997; Martiniello, 2008; Shaftel et al., 2006; Spanos et al., 1988)
Question wording/placing (Cummins et al., 1988)</td><td>Property of Numbers
Single Digit (Lean et al., 1990)
Multi Digit (Haghverdi et al., 2012)
Type of Number (e.g., fraction: (Raduan, 2010), decimal number: Problem Size, Role of Number (De Corte, 1988)
Number Magnitude (e.g., range of number smaller than 100: (Nesher, 1976)</td><td>Skills & Social Aspects
Social Consequences (Schley & Fujita, 2014)
Learning Disabilities (Kingsdorf & Krawec, 2014)
Successful/Unsuccessful Problem Solvers (Hegarty et al., 1995)
Calculation/WP difficulties (Powell & Fuchs, 2014)
Children/Adults (De Corte et al., 1990) / (Hegarty et al., 1992)</td></tr>
<tr><td>Required Operation
Addition Subtraction (De Corte & Verschaffel, 1987)
Multiplication Division (De Corte, 1988)
Given Number (Verschaffel & De Corte, 1990; Vicente et al., 2007)</td><td>Categorization
Semantic Structure of Arithmetic WPs (Riley et al., 1983)
Algebra Textbook Frequency (Mayer, 1981)
Standard /Non Standard WP (Jiménez & Verschaffel, 2014)</td></tr>
<tr><td rowspan="2">Semantics
Linguistics Verbal Cues (Van der Schoot et al., 2009)
Phrasing in Cue Words (LeBlanc & Weber-Russell, 1996)
Conceptual Rewording (Vicente et al., 2007)
Semantic/Object Relation (S. A. Martin & Bassok, 2005)
Presence of Distractor (Muth, 1992)</td><td>Mathematical solution strategy
Counting from larger number (De Corte & Verschaffel, 1987)
Position of the unknown (García et al., 2006)
Arithmetic fact retrieval (Orrantia et al., 2010)
Number Combination (Fuchs et al., 2009)
Situation/Mental Arithmetic Strategy (Brissiaud & Sander, 2010)</td><td>Solution Strategies
Algebra WP/Arithmetic WP (Koedinger & Nathan, 2004)
WP Solving Theory-Models (Kintsch & Greeno, 1985)
Translation Strategies (Hegarty et al., 1995)
Spatial/Visual representation (Boonen et al., 2013)
Situation Model (Thevenot et al., 2007)</td></tr>
<tr><td>Relevance of Information
(Terao et al., 2004)
Numerical Processes and Representation
(Göbel et al., 2014; MacGregor & Price, 1999)</td><td>Other Aspects
Pedagogical Factors (Lean et al., 1990)
Socio-Mathematics (Reusser, 1988)
Stereotypes thinking (Foong & Koay, 1997)
Real-Word Knowledge (Verschaffel et al., 1997)
Response Mode (De Corte, 1988)</td></tr>
<tr><td colspan="2">Consistency Effect (Lewis & Mayer, 1987)
e.g., revisited: (Pape, 2003)</td><td>Computer Tutors (Nathan et al., 1992)
Computer Simulation (Dellarosa, 1986)</td></tr>
</table>

Basic Linguistics Influence on Numerical Cognition (Lachmair, Dudschig, de la Vega, & Kaup, 2014)	
Working Memory (Swanson et al., 1993)	

Table 1 - 1 Selected Linguistic, Mathematical and General Factors Investigated in Previous Studies.

INFORMATION RELEVANCE AND STEP-WISE PROBLEM PROCESSING

One relatively extensively studied factor in WPs is the relevance of the information. Individuals have to extract the relevant information from the text in order to carry out the correct solution. Secondary information distracts people from recognizing the underlying mathematical relations (Schley & Fujita, 2014). This extra information may also be presented in the form of an extra number or an extra operational step – one-step (i.e.: one calculation step has to be performed) and two-step problems (i.e.: two calculation steps have to be performed). Problem complexity increases with the addition of steps (Nuerk et al., 2001; Terao et al., 2004), as well as the addition of irrelevant information to the problem (Kingsdorf & Krawec, 2014) Presence of extraneous information and the need for an extra step reduced the ACC of the students' solutions, because students believe that all of the numbers in a WP should be used. All other factors being kept constant, two-step problems are much more error-prone than one-step problems (Muth, 1992). However, it cannot be concluded that the reason for two-step problems being more difficult is arithmetic complexity, because in two-step problems, the WP has also become more difficult linguistically as it usually contains more phrases and semantic distractors. Additionally, the second problem is that a systematic variation is not always possible because in two step-WPs are also attached to specific calculations: addition is always with multiplication, subtraction always with division (Van der Schoot et al., 2009). For this reason, we have chosen one-step arithmetic WPs for the measurement but also here further investigation would be necessary between the experimental settings.

OTHER NUMERICAL PROCESSES AND REPRESENTATION

Several other numerical processes and representations have not been investigated in WPs. For instance, as shortly outlined above, one major factor in simple calculation problems, which can be studied in isolation, is the presence or absence of a carry operation. Children and adults take longer and commit more errors when computing the solution to a sum for which adding the units leads to a change in the number of tens (e.g., 14 + 9 = 23) (Deschuyteneer et al., 2005; Furst & Hitch, 2000) than when it does not (e.g., 11 + 12 = 23). This effect is known as the carry effect; in carry problems, a one needs to be carried from the unit slot to the decade slot. The carry effect is influenced by various processes, but even by language structure (Göbel et al., 2014). Language influences on the difficulty of the numerical computations within a WP have to our knowledge not been studied. Other central topics of numerical cognition such as, e.g., number and symbol sense contribute to WP solving are also open questions (MacGregor & Price, 1999). We have chosen some selected variables/factors, which have been investigated in the WP research.

CONNECTING LINGUISTIC AND MATHEMATICAL FACTORS

There are so many linguistic influences on numerical cognition and arithmetic that this justifies a special issue like this. For instance, number word structure seems to play an essential role. Children growing up with regular number word structure usually perform better in variety of numerical tasks from basic verbal counting up to arithmetic e.g.: Miller, Smith, Zhu, and Zhang (1995) or Dowker, Bala, and Lloyd (2008). In addition, the consistency of the order of the number word system and the Arabic number influences transcoding (Imbo, Vanden Bulcke, De Brauwer, & Fias; Pixner, Zuber, et al., 2011; Zuber, Pixner, Moeller, & Nuerk, 2009) number comparison (Klein et al., 2013; Moeller, Klein, Kucian, & Willmes, 2014; Nuerk et al., 2005; Pixner, Moeller, Hermanova, Nuerk, & Kaufmann, 2011) calculation (Brysbaert, Fias, & Noël, 1998; Colome, Laka, & Sebastián-Gallés, 2010; Göbel et al., 2014). In addition, reading direction influences numerical processes like the SNARC effect (Fischer & Shaki, 2014; Shaki, Fischer, & Petrusic, 2009); see Göbel, Shaki, and Fischer (2011) for reviews. Finally, grammatical and syntactic properties of elementary number words influence early number acquisition (Sarnecka, 2014) and spatial-numerical representations (Roettger & Domahs, 2015). The linguistic influence on numerical cognition is hardly debatable any more. In fact, Lachmair et al. (2014) argue for a connection of language and words, O'Neill, Pearce, and Pick (2004) states that the link between language and mathematics might originate from the same roots, and "required abilities are not that split up as we think.", and MacGregor and Price (1999) also argue that between language and mathematics in WPs there is deep connection: "that the cognitive ability that drives symbol processing is the connection between language and maths". Nevertheless, systematic variation of both linguistic and numerical/arithmetic factors in WPs are scarce – though Bassok et al. (1998) already found that semantic relations between objects in the text of mathematical WPs were highly positively correlated with arithmetic operations that took these objects as arguments. Neural correlates of visualization and verbalization during arithmetic WP study also suggest that mental arithmetic in WPs is influenced by language processing (Zarnhofer et al., 2013).

Some individuals find it more difficult to understand the text of a WP than to find its solution. However, children do not have a repertoire of "highly automatized schemata" for representing the different problem types (García et al., 2006), but learn to handle the operation of addition and subtraction, and understand numerical concepts before seeing WPs in the curricula (García et al., 2006). Therefore, it is not surprising that children make more errors when solving WPs compared to number problems (Geary, Hoard, Nugent, & Bailey, 2012; Koedinger & Nathan, 2004). Children are able to solve several types of addition and subtraction problems before they start formal schooling (De Corte & Verschaffel, 1987) and they already have problem-solving strategies when they get to school (Lean et al., 1990). Therefore, most of the studies implicitly assume that problem solvers always have the necessary basic arithmetic skills, even in the case of children. This may lead to the misconception that numbers may play a lesser role than they actually do and factors other than computational skills are a major source of difficulty with WPs (Nesher, 1976; Reusser, 1993). On the linguistic side, it is also important that difficulties in solving WPs have been reported that could be attributed neither to the lack of general reading comprehension skills (Hegarty et al., 1995) nor to the lack of general mathematical skills. Nevertheless, linguistics and numerical/arithmetic factors are usually not independently manipulated in WPs and not even dissociated by other means (e.g., regressions). What is more, their interaction is rarely studied. One example for systematically manipulation the interaction between numerical and linguistic features is, e.g., a study from Verschaffel and De Corte (1990).

LEXICAL CONSISTENCY EFFECT

One of the few frequently studied factors examining the relation between text and arithmetic problems is lexical inconsistency. Some WPs contain linguistic markers as "less" or "more". In the direct translation strategy (Hegarty et al., 1995) students simply associate "less" with subtraction and "more" with addition. They search for linguistic markers and keywords. In the problem model strategy, they construct a mental model of the problem and plan their solution on the basis of this model. Successful learners are more likely to employ the problem model strategy; they focus more on variables names and relational terms and successful problem solvers re-read the text less frequently (Pape, 2003) in the eye-tracking studies. Unsuccessful learners, on the other hand, seem to rely on the direct translation strategy ; they focus on numerals and on relational terms, and linguistics markedness in the (Hegarty et al., 1992) eye tracking study. This leads to wrong solutions in lexically inconsistent texts, where "more" is associated with subtraction and "less" with addition. To give an example for lexical inconsistency, consider the following WP adapted from Boonen et al. (2013): “At the grocery store, a bottle of olive oil costs 7 euro. That is 2 euro more than at the supermarket. How much will [a bottle of olive oil] cost in the supermarket?” The anticipated difficulty in comprehension and finding the correct solution is due to the fact that the adverb “more” evokes the concept of addition, but the correct solution is not 7+2 but 7-2, given the way the text is organized. Verschaffel et al. (1992) found such a reaction time consistency effect for children but not for adults. Nesher (1976) and Lean et al. (1990) obtained similar results in experiments with groups of non-disadvantaged children and students, showing that linguistic semantic consistency with respect to the required mathematical operation is an important determinant of task difficulty. Inconsistent language results in a high error rate and longer response time (Hegarty et al., 1992), even in Verschaffel (1994) retelling one-step compared WPs showed a strong evidence for the consistency hypothesis. Students made approximately 13% more reversal errors on inconsistent than on consistent language problems and the difficulty of comprehending inconsistent-language problems were increased when the correct arithmetic operation was an increase. However, the literature is inconsistent if the consistency effect is present in both students and children. Children find it easier to convert the relation term “more than” into subtraction operation than the relational term “less than” into an addition operation (Lewis & Mayer, 1987; Pape, 2003; Van der Schoot et al., 2009; Verschaffel et al., 1992).

When neither reading comprehension nor arithmetic skills alone can explain failure to solve WPs, a possible explanation is that linguistic complexity and numerical complexity rely on the same resources (e.g., working memory). The premise is that there is not an absolute atomic concept of difficulty for WPs. Rather; there are multiple linguistic and numerical/arithmetic factors which contribute to a problem’s complexity. It is a combination of these factors that might make a problem additively more or less difficult because they exert demands on more general resources like working memory. Generally, problem-solving performance is related to the ability of reducing the accessibility of no target and irrelevant information in the memory (Passolunghi & Siegel, 2001). Working memory contributes to early arithmetic performance, and studies also show that this extends to WP-solving (Lee, Ng, Ng, & Lim, 2004) due to semantic memory representation “less than” which is more complex than “more than”. Changes in the structure of the text makes more demand on the working. It has been suggested that WPs in general are related to working memory (Swanson et al., 1993). This will probably also be influenced by instruction specifying how participants have to solve a WP, and the method of evaluation and scoring system. In Kintsch’s model (Van Dijk & Kintsch, 1983) of reading comprehension, working memory is used to keep a number of text propositions active simultaneously. In particular, working memory has been related to each single component mentioned above, such as text-problem relation, the linguistic complexity, and the arithmetic complexity.

FUTURE DIRECTION, OPEN QUESTIONS

WP difficulty is influenced by the complexity of linguistic factors, numerical/arithmetic factors and their interrelation. To better understand the difficulty of WPs, it would be desirable to manipulate such variables and their interaction following the principle of isolated variation. To support a systematic investigation, the variables to be manipulated also need to be discussed against the backdrop of the relevant conceptual and empirical issues in the underlying fields, linguistics and numerical cognition. This has too rarely been the case in the past. For instance, in the earlier studies on algebra WPs, the linguistic cues are of mixed categories (adverbs, verbs, nouns, etc.) and the effect of the complexity of syntactic structures is not considered. Similarly, numerical complexity like basic number properties (e.g., magnitude, place-value processing for multi-digit numbers) or the complexity of underlying arithmetic computations (e.g., carry effects for addition, relatedness or consistency effects for multiplication) are often neglected. WP research would be well advised to consider the foundational categories, properties and findings of both numerical cognition and linguistics when it examines which WPs are difficult for which groups and why. Not only the main effects of numerical and linguistic complexity should be studied, but also their interaction. To make the relevant aspects explicit, Figure 1 sketches an overall process model of WP-solving.

The joint investigation of linguistic and numerical processes also needs to consider joint moderator variables such as working memory in order to explore the possible interactions between them. Since working memory affects all components of complexity of a WP, the difficulties triggered may not be simply additive, but interactive. The resolution of linguistic and numerical difficulties may rely on the same processing stages and resources (Sternberg, 1969). To investigate this, more collaboration between linguists and numerical cognition researchers would be desirable.

Finally, we suggest a differential-psychological approach to WP research. Different students may have a problem with different types of WPs. Linguistically rather weak students may have problems with linguistically complex WPs, and arithmetically rather weak students with arithmetically complex problems. Undifferentiated presentation of WPs in experiments will not provide sufficient information about which skills and processes an individual child should practice.

Only with such differentiation on an item level (as regards linguistic and numerical complexity and their interrelation) and on an individual level (as regards linguistic and numerical skills and general cognitive abilities) will it be possible to understand why a particular child has its individual difficulties with particular WP types. Such an understanding, however, is essential to promote tailored learning of one of the most difficult arithmetic problem types that students encounter in school.

Acknowledgement

This research was funded by the LEAD Graduate School [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Magdalena Wolska is a Junior Research Group Leader of the LEAD Graduate School and her work is supported by the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft, ZUK 63). Gabriella Daroczy is a doctoral student of the LEAD Graduate School.

STUDY 2: THE INTERACTION OF LINGUISTIC AND ARITHMETIC FACTORS AFFECTS ADULT PERFORMANCE ON ARITHMETIC WORD PROBLEMS

Contributions of Co-Authors and other Persons

Authors

Gabriella Daroczy[1,3], Detmar Meurers[2,3], Jürgen Heller[1], Magdalena Wolska[3], & Hans-Christoph Nürk[1,3]

[1]Department of Psychology, Eberhard Karls University Tuebingen, Tuebingen, Germany
[2]Department of Linguistics, Eberhard Karls University Tuebingen, Tuebingen, Germany
[3]LEAD Graduate School & Research Network, Eberhard Karls University Tuebingen, Tuebingen, Germany

Author	**Author position**	**Scientific idea (%)**	**Data generation (%)**	**Analysis and Interpretation (%)**	**Paper writing (%)**
Daroczy, G.	1	60	100[1]	60	70
Meurers, W. D.,	2	15	0	5	10
Heller, J.,	3	0	0	20	5
Wolska, M.,	4	10	0	5	5
Nuerk, H.-C.	5	15	0	10	10

Data generated in the course of student work

Vesna Mikavica, Bachelor-Thesis, Title: *"Mathematische und sprachliche Faktoren in Textaufgaben"*, 2014 SS

Proof-Reading

Juliane Skinner, Zoë-Lauren Kirste

I/We certify that the above-stated is correct.

28.09.2020

Date, Signature of at least of one of the supervisors

[1] Responsible, Vesna Mikavica helped in data collection (50%)

ABSTRACT

Performance on word problems is influenced by linguistic and arithmetic factors, and by their interaction. To study these factors and interactions, we manipulated linguistic and arithmetic factors independently in a within-participant design that included complexity parameters a) in the domain of arithmetic: carry/borrow (no-carry/borrow vs. carry/borrow), operation (addition vs. subtraction), b) in the domain of linguistics: nominalization (nominalized vs. verbalized form), and c) linking the two domains: lexical consistency (linguistic predicate locally consistent vs. inconsistent with mathematical operation). Response times of 25 students solving 320 one-step word problems were measured. All four factors showed a main effect on response times, and interactions between linguistic and arithmetic factors affected response times. These interactions were observed when the linguistic and arithmetic factors were conceptually linked. Our results highlight that not only the linguistic and arithmetic complexities of an item contribute to the difficulty of a word problem, but linguistic and arithmetic factors interact. We discuss the theoretical implications for the numerical and the linguistic domain as well as the possible impact of domain-general characteristics, such as working memory limitations as a potential reason for the observed interactions between numerical and linguistic attributes.

Keywords

arithmetic word problem, nominalization, carry effect, lexical consistency, addition, subtraction, linguistic complexity

INTRODUCTION

Word problems, where a mathematical problem is presented using language, play an important role in the school curriculum, and they have been shown to be difficult for people of all ages (Boonen et al., 2013; Hegarty et al., 1992; Lewis & Mayer, 1987; Nesher & Teubal, 1975; Riley, 1984; Verschaffel et al., 1992). A wide range of factors, from individual traits to the socio-economic background, the school environment, and the scoring system influence performance on WPs. Here we focus on the fact that the ability to successfully solve a WP depends on task characteristics: how the text is formulated (i.e., linguistic features), and how difficult the arithmetic operations are (i.e., arithmetic features). The connection between linguistic and arithmetic factors is not yet fully understood. One reason is that linguistic and arithmetic factors are confounded in many studies, making it hard to draw conclusions about which of the factors – i.e., the arithmetic or the linguistic one – are responsible for increased response times or higher error rates. For example, experiments with WPs often differ in the number of computational steps: many studies compare performance on one-step problems, involving one calculation, to two-step WPs, which require two calculations involving multiplication or division in combination with addition or subtraction. Two-step problems are usually arithmetically and linguistically more complex, containing longer sentences with more propositions. Generally, they also are more complex in terms of domain-general factors given that for two-step problems more information must be kept in working memory (Fuchs, Gilbert, et al., 2016). Thus, when the additional step in two-step problems impacts the speed and/or the ACC of solving the task (Muth, 1992; Quintero, 1983), it is not clear whether the increased difficulty is due to arithmetic complexity, linguistic complexity, or the domain-general requirement of more cognitive resources.

In this study we spell out and investigate a distinction, that has not been prominent so far, between 1) linguistic and numerical/arithmetic factors that are conceptually independent of one another, in the sense that they can also be manipulated independently, and 2) those which are intrinsically linked. Indeed, some linguistic and arithmetic factors are not separable. For example, a keyword that provides a hint for the solution of a WP is a linguistic factor that can only be chosen based on the mathematical operation necessary to solve the text problem (e.g., Van der Schoot et al., 2009). Similarly, the arithmetic factor operation is directly connected to the lexical choice as a linguistic feature of WPs (Lave, 1992), as the concepts of addition and subtraction are commonly expressed by “giving” and “taking” in everyday language (Terezinha N Carraher, Schliemann, & Carraher, 1988). On the other hand, there are independent linguistic and arithmetic factors, where the other domain remains unaffected. For example, changing the name of the protagonist in a WP (Abedi & Lord, 2001) does not generally change the underlying mathematics. Similarly, whether the numbers in a WP are two-digit or three-digit numbers (Thevenot & Oakhill, 2006) does not change the language. Such intrinsically related and unrelated factors may affect the process of solving a WP to different extents. At the same time, linguistic complexity and arithmetic complexity in WPs are often seen as subsequent additive processes, which do not interact with each other. If such a theoretical assumption were true, then one could argue that it is enough to study arithmetic complexity in common (nonverbal) arithmetic tasks. However, studying arithmetic complexity in isolation would not make it possible to observe and distinguish different sources of variance in solving WPs and establish whether they arise from independent linguistic factors, independent arithmetic factors, or factors involving both domains. It is also impossible to pose the general question as to whether all of these factors may arise from domain-general requirements for cognitive resources. Therefore, in our view it is necessary to explore the manipulation of both linguistic and arithmetic factors in a WP study independently.

In the sections Linguistic Complexity and Arithmetic Complexity, we will introduce the related and unrelated factors. Following that, in the section Underlying Cognitive Processes of Arithmetic Problem-solving we give an overview of existing models for solving WPs. Finally, in the section Interaction of Linguistic and Arithmetic Factors, we will first elaborate on how the presence or absence of an interaction between the factors could support the existing models, before closing with a discussion of potential interactions between unrelated and related arithmetic and lexical complexity.

Linguistic Complexity

Performance on one-step problems is strongly dependent on the wording of the problem (De Corte et al., 1985; Vicente et al., 2007). Some linguistic factors affect the underlying mathematical structure, while other factors do not. Factors not affecting the underlying mathematical structure include general descriptive characteristics (e.g., overall number of words, average sentence length), grammatical features, and most lexical properties (Haag, Heppt, Stanat, Kuhl, & Pant, 2013; Richards & Schmidt, 2013). Factors that affect the mathematical structure, like the consistency of the key word expressing the operation (Domahs et al., 2006; Pape, 2003; Van der Schoot et al., 2009; Verschaffel et al., 1992) contribute to the difficulty of WPs. The relation of these factors to the problem-solving process is often not clear.

Linguistic Factors Unrelated to Arithmetic Factors

To begin with, in some cases modification of the language of a WP without changing the semantic and mathematical structure leads to higher success rates (Cummins et al., 1988; Davis-Dorsey et al., 1991; Stern & Lehrndorfer, 1992; Vicente et al., 2007). However, it is important to note that richer text can also bring about additional difficulty, depending on the linguistic complexity of the added text. Adding non-relevant information (Barbu & Beal, 2010), a modification that does not change the mathematical structure of the text, can lower success rates. Nevertheless, many such text manipulations unrelated to arithmetic factors are reported to improve solution ACC:

(i) The text is described more richly to clarify the situation (Stern & Lehrndorfer, 1992). Consider the example from Hudson (1983): without situational rewording "Here are some birds and here are some worms. How many more birds than worms are there?' and with situational rewording "Here are some birds and here are some worms. Suppose the birds all race over and each one tries to get a worm. How many birds won't get a worm?"

(ii) Conceptual rewording where the underlying structure is highlighted through simple language modification (e.g., Vicente, Orrantia, & Verschaffel, 2008).

(iii) Introduction of personalization (Abedi & Lord, 2001; Reusser, 1988) to the context.

It is also suggested that uncommon vocabulary in mathematical tasks is negatively related to performance (Abedi & Lord, 2001; Shaftel et al., 2006). However, other studies such as that of Bergqvist, Dyrvold, and Österholm (2012) did not find any evidence for this relation.

In choosing one linguistic factor that does not change the underlying mathematics but increases the complexity of the language, we decided on *nominalization*. *Nominalization* is the process of turning verbs, often expressing actions, into nouns (Francis, 1989), thus increasing the number of noun phrases in the text. For instance, the verb *"to earn" can be nominalized into "the earning",* as illustrated by the following WP examples, with the verbal form in (1) and the nominal form in (2) shown in bold (In German, the language we are focusing on in the experiments, nominalized forms are systematically used and typical for academic language, while in English such nominalization can be marginal):

(1) *A man saved some money. He had 82 Euros.*
The next day, he ***earns*** *15 Euros.*
How much money does the man have now?

(2) *A man saved some money. He had 82 Euros.*

The next day, he is happy about the ***earning*** *of 15 Euros.*

How much money does the man have now?

These examples differ in terms of their linguistic complexity, but both problems can be solved with the same arithmetic operation, in this case addition. It is known that nominalization increases the difficulty of comprehension (Michael Halliday et al., 2014; To et al., 2013) independent of the domain of WPs. The number of noun phrases has thus been referred to as the foremost predictor of text difficulty (Haag et al., 2013). In addition, nominal style is characteristic of academic language, in legal, political, or scientific texts (Baratta, 2010). It is also common in mathematical discourse (Perry, Howard, & Miller, 1999). Nevertheless, nominalization has rarely been investigated in WPs. In one study that investigated nominalization it was shown that if WPs contain a high density of nominalization this significantly affects the solution rate for 10^{th} grade students (Schlager et al., 2017). However, in the study, multi-step and more complex problems were used, manipulating various other linguistic features at the same time.

While nominalization does not change the underlying mathematical structure and does not introduce additional arithmetic complexity into identifying the required calculation, we expect that nominalization, as a characteristic of complex language, should affect WP performance.

LINGUISTIC FACTORS RELATED TO ARITHMETIC FACTORS

Complementing the unrelated linguistic factors just discussed, there also are linguistic factors that are related to the arithmetic underlying a WP. The most prominent of such factors is *lexical consistency*, introduced in the work of Lewis and Mayer (1987). Lexical consistency concerns specific keywords in the text, the so-called cue words, which signal or hint towards particular arithmetic operations like addition or subtraction (Hinsley et al., 1977). A WP is considered lexically consistent if the semantics of the cue words signal an operation that is congruent with the operation required for the correct solution. It can be illustrated by the following example:

(3) *A man saves money on some purchases. He had 82 euros. He* ***earns*** *15 euros. How much money does the man have now?*

Here the cue word *earns* is associated with getting more, i.e., it infers the operation can be associated, in this case, with addition, and the solution indeed requires addition: 82+15. In contrast, in inconsistent WPs, the relational term hints at some other operation, usually the opposite.

(4) *"A man saves money on some purchases. He* ***spent*** *74 euros. He has 23 euros now. How much money did the man have?"*

In example (4) the cue word *spent* is commonly associated with having less euros than before; thus, it is associated with subtraction. However, in this example, computing the solution requires addition: 74+23 = 97. Such problems are called lexically inconsistent because the operation most frequently associated with the lexical meaning of the cue words is inconsistent with the operation needed for the correct solution in this particular context. Indeed, most errors in lexically inconsistent conditions are due to an erroneous choice of operation (Hegarty et al., 1992; Zawaiza & Gerber, 1993). In this manuscript, we refer to lexical consistency as a linguistic factor *related* to arithmetic factors, because the word chosen is linked to the mathematical operation. Note, however, that it can be manipulated in a 2 x 2 design (lexical consistency x operation) since the operation to be carried out ultimately depends not on the most common context of use of a particular trigger word, but on the particular question asked by the given WP. Lexically consistent and inconsistent problems can be constructed for both addition and subtraction, by combining the appropriate cue words for the operations with different question contexts. In the current study, we have done so.

The effect of lexical consistency on both solution time and ACC has been studied previously (e.g., Hegarty et al., 1992; Hegarty et al., 1995; Pape, 2003; Verschaffel et al., 1992). Compared to lexically consistent problems, lexically inconsistent ones are usually associated with increased solution time and lower solution ACC, although some studies failed to detect effects. For example, Verschaffel et al. (1992) found a consistency effect for children but not for adults when they were presented one-step arithmetic comparison problems. The studies of Hegarty et al. (1992) and Hegarty et al. (1995) detected the consistency effect in two-step arithmetic problems and showed that even students who successfully solve WPs need more time to derive solutions in the inconsistent condition. In line with this, we expect lexical consistency to affect WP performance.

ARITHMETIC COMPLEXITY

Naturally, the complexity of a WP also depends on arithmetic factors. For instance, choosing the correct operation strongly depends on the nature of the numbers in the problem (De Corte et al., 1990), and more difficult calculations might lead to lower solution ACC (Thevenot & Oakhill, 2006). For example, participants may show more calculation failures when operations requiring a carry are involved. The arithmetic factors can be grouped according to their relatedness to linguistic factors.

ARITHMETIC FACTORS UNRELATED TO LINGUISTIC FACTORS

It is well known from the literature that arithmetic complexity in the form of basic number properties (e.g., Ashcraft, 1992; Nuerk et al., 2011) and the complexity of underlying arithmetic computations (e.g., Göbel et al., 2014) play a major role in common arithmetic problems. Surprisingly, there has been relatively little effort to investigate such basic forms of arithmetic complexity in WPs, although it is known that some errors originate from arithmetic computation errors themselves (Kingsdorf & Krawec, 2014; Raduan, 2010). Multiple effects of numerical cognition are known to contribute to the difficulty of an arithmetic problem (see, Nuerk et al. (2011) for multi-digit numbers). We decided to start with arguably the most ubiquitous effect of multi-digit addition complexity in arithmetic cognition, the carry and borrow effect. The difficulty of two-digit addition and subtraction increases whenever a carry or borrow operation is required (Artemenko et al., 2018). In carry problems, a 1 needs to be carried from the unit slot to the tenths slot. An example for arithmetic calculation with carry is 14 + 39 = 53, where 4 + 9= 13 and the digit 1 is the carry; an arithmetic calculation with no-carry is illustrated by 11 + 12 = 23. Similarly, the borrow operation in subtraction is needed whenever the unit of the minuend is smaller than the unit of the subtrahend so that a decade must be borrowed from the minuend. Examples for arithmetic calculation with borrow is 34 – 19 = 15 where 4 - 9= -5 so we compute $(10 - 9) + 4 = 5$ and the 10 is obtained by taking ("borrowing") 1 from the next digit to the left. In contrast, 35-12=23 is an example for an arithmetic calculation with non-borrow. In sum, carrying and borrowing increases the arithmetic complexity of a problem. Importantly, the carry or borrow operation is unrelated to any linguistic property. Merely the numbers are changed, the problem is otherwise identical. It has been shown that in carry/borrow conditions compared to non-carry/non-borrow conditions the response time increases for two-digit addition in children and adults (Artemenko et al., 2018).

In sum, there is reliable evidence outside of the WP literature that the carry effect influences the arithmetic performance of both children and adults and evidence that other arithmetic effects also play a role in WPs. Therefore, we hypothesize that the carry effect should also influence performance on WPs, but this has – to the best of our knowledge – never been systematically tested.

ARITHMETIC FACTORS RELATED TO LINGUISTIC FACTORS

One of the major challenges in a WP is to find the correct arithmetic operation after understanding the text. Nevertheless, even after the correct operation is successfully detected, the operations might still vary in their difficulty. Most WP research addresses addition and/or subtraction as the operation (e.g., Carpenter et al., 1981; De Corte & Verschaffel, 1987; Fennema et al., 1996; García et al., 2006; Klein et al., 2009; Moeller, Klein, et al., 2011b). Presumably, this is because the limits of mental arithmetic are quickly reached in more complex operations, such as multiplication or division (Swanson, 2004).

In our experiments, we also focus the WP operation tasks on addition and subtraction. Incorporating these two operations is necessary. If the same operation needs to be computed for all problems, e.g., addition, people will not try to read and understand the text anymore, but simply add the numbers found in the text, no matter what the text says.

With regard to the effect of the operation itself, it was shown that subtraction is generally more difficult (Artemenko et al., 2015) and elicits greater response times (Orrantia et al., 2012) than addition. Additionally, many types of subtraction tasks can be solved by various strategies, including the indirect addition or subtraction by addition strategy (De Corte & Verschaffel, 1987; Torbeyns et al., 2009). Therefore, we expect that subtraction should increase the time needed to solve a WP, as compared to addition.

Concluding, we sketched arguments for why each of the above mentioned linguistic and arithmetic factors could affect performance on WPs. In order to differentiate the extent to which they influence performance and to explore their interactions we therefore divided task characteristics into unrelated and related factors. Additionally, so far there is very little information about the relationship of these factors to problem-solving processes. Differentiating the related and unrelated factors may also be important because different WP-solving models would suggest different interactions between those factors – which we discuss in the following section.

Underlying Cognitive Processes of Arithmetic Problem-solving

Models of Problem-solving

Several studies on WPs investigated the underlying cognitive processing (De Corte et al., 1990; Hegarty et al., 1992; Hegarty et al., 1995; Verschaffel et al., 1992). It is hypothesised that solving WPs requires four distinct phases (Mayer, 1984): an initial reading phase (i.e., translation of the text), integration (i.e., mental representation and the construction of the problem model), planning (i.e., generating a solution plan), and solution execution (i.e., calculation). However, a comprehensive theory of problem-solving is still lacking (Passolunghi & Pazzaglia, 2004). The existing models for solving WPs differ in several respects, such as in the nature and the origin of the internal problem model — the mathematization of the text — and whether the problem-solving phases are fully separable or not. First, it is under debate whether the internal problem representation is a result of schemas, a situated model or a mental representation model. According to Kintsch and Greeno (1985), the problem model relies only on schemas stored in long-term memory and the problem model is constructed from the text base and relies on the problem-solver's previous knowledge as well as the text. Van Dijk and Kintsch (1983) extended this prior schema-based problem model construction with a situation model, which corresponds to a level of representation that specifies the agents, actions, and relationships between events in everyday contexts. Finally, some argue that ad hoc transient mental representations are constructed for each problem encountered with the help of working memory (e.g., Thevenot & Barrouillet, 2015; Thevenot et al., 2007). Finally, the problem-solving models in the literature do not agree on whether the process of reading is completely separable from the process of solving. Studying the presence or absence of an interaction between related and unrelated factors from the linguistic and arithmetic domain could provide relevant evidence for this debate.

Interactions between Related and Unrelated Factors and the Models of Problem-solving

In the first model, the propositional theory, the text formulation determines the difficulty of constructing an adequate representation (Abedi & Lord, 2001; Cummins et al., 1988; De Corte et al., 1985). This is supported by the documented strong connection between text comprehension and the solving of WPs (Boonen et al., 2013; Boonen et al., 2014; Kintsch & Greeno, 1985). This model sees the phases of problem-solving as clearly separable and serial. For example, in the case of an arithmetic WP, the complexity of the text rather than the mathematical operations involved influences the processing of the problem (Nesher, 1976). This is supported by Rabinowitz and Wooley (1995), who found no significant interaction between problem size (one digit vs. two digit numbers) and various problem types. According to the propositional model the number difficulty (i.e., carry) should mainly affect the calculation/execution phase, and text difficulty (i.e., nominalization) mainly the initial text comprehension phase, and not the building of the problem model phase. Contrary to this, factors that are related to other arithmetic/linguistic factors should affect not only the first reading (in the case of lexical consistency, which has been shown to manifest itself in the second phase of problem-solving) and the calculation (i.e., in the case of operation), but also the phase where the problem model is built. This is also supported by the fact that an interaction between operation and lexical consistency is often found in studies (e.g., Van der Schoot et al., 2009; Verschaffel et al., 1992).

On the other hand, other models suggest that the text formulation affects not only the first reading but also other phases in the solution process, and that problem comprehension and computational processes interact. Such an interaction could cause the computational requirements of the problem to interfere with problem representation. For instance, Thevenot and Oakhill (2005) suggested that the mental representation constructed when solving a WP involves a re-enactment of the solvers' experiences with the processing of magnitude information for quantities. The study of Munez et al. (2013) also suggests that in solving an arithmetic WP, solvers construct a magnitude-based mental representation that goes beyond a conceptual representation in the form of propositions. This would suggest that number processing might also influence another stage of problem-solving. This hypothesis is supported by De Corte et al. (1990), who questioned the sequential and linear character of a theoretical model of competent problem-solving, especially with respect to more complex problem types. The theoretical model proposed by Bergqvist and Österholm (2010) also indicates that the process of solving WPs comprises a cyclical component revisiting the mental representation. This would also mean that reading interacts with other phases of problem-solving. These non-sequential models would suggest an interaction, for example, between the unrelated arithmetic factors/linguistic factors and the other factors. This hypothesis is supported, for example, by the findings of Hegarty et al. (1992) or Verschaffel et al. (1992), who found that in WPs with linguistically marked words ("less than") more time is needed to solve the problem compared to WPs with unmarked words ("more than"). Markedness suggests that in most languages, there is usually a complementary pair of adjectives, with one adjective being the ground (unmarked) form and the other being the derived opposite (marked) form. Which adjective is marked or not can be determined in three ways: The easiest and most consistent way is formal markedness. In this case, the form of the marked adjective is explicitly marked by a negating prefix, for example the prefix "in" turning "efficient" into "inefficient", or the prefix "dis" that can mark "organized" to obtain "disorganized" (Zimmer, Austerlitz, Diver, & Martinet, 1964). Two other ways to define which is the marked term are semantic and distributive markedness (Lyons, 1977). An example of semantic markedness is "old" vs. "young", where in neutral context the unmarked term ("old") is the one used ("How old is the baby?"). Distributive markedness, on the other hand, is related to word frequency.

In conclusion, each model predicts different interactions. According to the propositional theory, we should find no interaction between related linguistic factors/arithmetic factors and unrelated arithmetic factors/linguistic factors given that this theory sees the initial reading, the calculation phase, and the building of the problem model phase as distinct, non-overlapping stages of problem-solving. Applying the logic of Sternberg (1969), interactions are not possible because there is no common stage of processing for the cognitive processes underlying the manipulated factors. In contrast to propositional models, we should find an interaction between related and unrelated factors according to the models that consider problem-solving to be sequential and cyclic because reading and number processing might both influence the problem model phase, i.e., the manipulated factors operate partially on a common stage of processing. In our case, we can associate the unrelated mathematical factor carry mostly with the calculation phase as it increases the difficulty of the calculation. On the contrary, the unrelated factor nominalization should be associated with the initial reading phase, because it increases the reading demand only. The factors lexical consistency and operation could be associated with the phases where the mental representation is constructed, as they are related to both linguistic and mathematical domains. This assumption is consistent with the literature suggesting both factors affect the second stage of problem-solving (e.g., Hegarty et al., 1992). This indicates that an interaction between the unrelated carry factor and any other factors would suggest that the calculation process interacts with other problem-solving phases. Similarly, an interaction between the nominalization factor and any other factors would suggest that the reading comprehension phrase interacts with other problem-solving phases.

Nevertheless, models for solving WPs usually do not consider the joint investigation of numerical and textual difficulty, and their involvement in mental problem-solving. For example, the role of numerical information is especially unclear and often not covered in the models. Therefore, in the next section, we will elaborate on the possible interactions of related and unrelated linguistic and mathematical task characteristics.

INTERACTION OF LINGUISTIC AND ARITHMETIC FACTORS

Linguistic and arithmetic factors may influence the problem-solving processes differentially. For instance, a focus on specific parts of WPs can be associated with certain problem-solving strategies. Expressed in terms of the strategies characterized by Hegarty et al. (1992), students using a so-called direct translation strategy, where students select keywords and numbers from the text to carry out a computation on this shallow basis, would focus on a few words only, whereas students using a problem-model strategy would pay more attention to the a broader range of words in the problem and build a situation model. However, under semantically less demanding conditions problem solvers can apply successful strategies (Van der Schoot et al., 2009). In addition, looking at the keywords without an understanding of the problem situation does not necessarily lead to a superficial solution process. As the difficulty of the text increases, the creation of the situation model and analytic processes operating on the selected items to generate inferences (Evans, 1984) get more and more important.

According to the model of Daroczy, Wolska, Meurers, and Nuerk (2015), WP difficulty comprises linguistic factors and arithmetic factors and these affect individual performance both directly and through mediator variables, such as domain-general attributes and solution strategies. Due to the joint resources, interactions with linguistic factors are to be expected and task characteristics should not affect the solution phases to the same extent. The interaction of linguistic and arithmetic factors should be present in the response time because some factors are not separable from the other domain and play an extensive role in building mental models and we hypothesise that these factors should not affect the solution phases to the same extent. If the factors are related by their very definition there must be a direct connection for such related linguistic and cognitive factors given that the relation does not only appear when working memory or other domain-general resources are relevant. For example, subtraction is more heavily dependent on mental representation than a simple numerical manipulation, because different semantic classes correspond to the different types of conceptual knowledge needed to solve problems, such as knowledge about increases or decreases in quantity. The interaction with the related factor (lexical consistency, operation) should therefore be especially pronounced because they all overlap in the reading, comprehension and execution phases. We expect this interaction to be over-additive, i.e., when the relation between operation and lexical consistency is such that both factors are difficult (lexically inconsistent subtraction problems) – due to limited resources – and we hypothesize longer response times than expected based on the main effects only.

However, according to the model of Daroczy et al. (2015), we would expect interactions not only between related factors but also between unrelated and other factors. The complex linguistic and mathematical task characteristics make the task more difficult (leading to main effects in appropriate designs), but they can increase domain-general attributes such as cognitive load because they impose on limited domain-general resources (Sweller, 1994), and because they might share the same processing stage (e.g., Sternberg, 1969). However, the mechanisms of and relationship to other domain specific and domain general factors are still under debate (Lee, Ng, & Ng, 2009; Tolar et al., 2012; Zheng, Swanson, & Marcoulides, 2011). Several studies have shown that both domain general factors, such as working memory (Adams & Hitch, 1997; Passolunghi & Siegel, 2001; Swanson, 2004), reading comprehension (Swanson & Beebe-Frankenberger, 2004), and processing speed (Kail & Hall, 1994) as well as domain specific factors, such as arithmetic computing or concept formation (Fuchs, Gilbert, et al., 2016), are related to WP performance. Furthermore, for other domain specific and domain general factors, working memory was shown to play a role not only for children (e.g., the verbal and spatial components of working memory (Soltanlou et al., 2015) but also for adults. For instance, for adults, the central executive is consistently found to be important for the carry operation in multi-digit arithmetic, e.g., Imbo, Vandierendonck, and De Rammelaere (2007), and in the domain of WPs working memory capacity influences the choice of solution strategies (Thevenot & Oakhill, 2006). Working memory also plays a role in language acquisition and understanding prepositions (N. C. Ellis, 1996). Its limitations affect the ability of elderly adults to process complex syntactic constructions (Norman, Kemper, & Kynette, 1992). This means, for example, for factors unrelated to arithmetic/linguistic, such as carry and nominalization, we hypothesize that the interactions should still be observed but less consistently. We wish to note that such an interaction does not need to rely on a direct relationship between linguistic and arithmetic factors per se, for unrelated factors like nominalization an interaction may simply entail linguistic and arithmetic factors using the same type of domain-general resource at some stage of the solution process.

OBJECTIVES

In this paper, we examine the role of linguistic and arithmetic factors in WP-solving performance. In designing the items, we manipulated linguistic complexity independent and orthogonal to the arithmetic complexity of the arithmetic problem underlying the WP. We then designed the study such that for both linguistic and arithmetic complexity, there was one factor relating linguistic and arithmetic complexity and one unrelated factor. In particular, the linguistic factor lexical consistency relates to arithmetic complexity (namely operation), whereas the linguistic factor nominalization does not. Analogously, the arithmetic factor operation is related to linguistic complexity (namely lexical consistency), whereas the arithmetic factor carry/borrowing is not. Based on the literature reviewed above, we formulated the following hypotheses:

1. We hypothesize that all linguistic and arithmetic factors will show a main effect on performance (H1), which we have split up for each factor as follows:
    - Regarding arithmetic, subtraction tasks should take significantly longer than WPs with addition (H1.1).
    - Because of the carry effect, WPs with a carry operation will result in significantly longer response times (H1.2).
    - The consistency effect will cause significantly longer response times for lexically inconsistent items (H1.3).
    - Additionally, we assume that the nominal form is more difficult to understand even for adults and will increase the response time significantly (H1.4).
2. Interactions between linguistic and arithmetic factors are expected (H2):
    - For linguistic/arithmetic factors which are related to other arithmetic/linguistic factors, such interactions should be particularly pronounced and more consistent, i.e., in the case of lexical consistency and operation (H2.1).
    - For factors affecting only linguistic or arithmetic, i.e., carry and nominalization, the interactions should still be observed, but less consistently, i.e., in some cases they will be absent. (H2.2).

METHODS

PARTICIPANTS

A total of 29 students participated in the experiment. All were native German speakers between 18 and 45 years old with normal or corrected (only with soft contact lenses) to normal vision. Neurological or psychological disorders were exclusion criteria. Four persons were excluded from the analysis for the following reasons: In two cases, the experiment had to be stopped due to technical difficulties. The other two subjects had an error rates that were too high (participants with an ACC below 75% were removed). Twenty-one of the remaining 25 persons ($M = 22.08$, $SD = 2.59$) were female ($M = 21.86$, $SD = 2.23$) and four were male ($M = 23.25$, $SD = 3.77$). For the remaining participants, the error range was from 1.0% to 13.3% ($M = 7.0\%$; $SD = 3.4\%$).

Participation was on a voluntary basis and was rewarded with either three subject hours towards course credits or with 20 Euros. Informed consent was given by all participants. The study was performed in accordance with the ethical standards of the Declaration of Helsinki.

STIMULI AND DESIGN

The study was a 2x2x2x2 design with the main factors: operation (addition/subtraction), carry (carry/non-carry), lexical consistency (consistent/inconsistent form), and nominalization. 320 simple arithmetic, one-step WPs in German were designed for the study. There were 16 conditions that consisted of 20 sentences (see Table 1 and Table 2). The four factors were manipulated simultaneously and orthogonally. Specifically, to explore the effect of arithmetic complexity, the numbers and operations were manipulated, while the text remained largely identical. Likewise, for exploring linguistic complexity, linguistic factors were manipulated, while the complexity of the arithmetic problem was kept constant. All sentences belonged to the type "Change" – namely Change 1, Change 2, Change 5 and Change 6 – according to the categorization of Riley (1984). Change problems refer to dynamic situations in which some event changes the value of a quantity (Verschaffel & De Corte, 1993). All four types of these mentioned problems were equally distributed. In half of the stimuli the result set (160 tasks) was unknown, from which half belonged to the category Change 1 (80 tasks, addition), the other half to the category Change 2 (80 tasks, subtraction). From the other 160 problems – where the start set was unknown – half belonged to the category Change 5 (80 tasks, addition), and the other half Change 6 (80 tasks, subtraction). Change 1 and Change 2 are considered to be easier tasks than the more difficult Change 5 and Change 6 (Riley, 1984) which means that both addition and subtraction tasks contain one easier and one more difficult type.

Table 2 - 1 *16 Conditions.*

	Consistent Form		Inconsistent Form	
	Nominal Form	Verbal Form	Nominal Form	Verbal Form
Subtraction				
Carry/Borrow	C1	C2	C3	C4
Non-Carry/Non-Borrow	C5	C6	C7	C8
Addition				
Carry/Borrow	C9	C10	C11	C12

Non-Carry/Non-Borrow	C13	C14	C15	C16

Table 2 - 2 *Examples for the 16 Conditions.*

Condition	Sentences
C1	Ein Mann spart Geld für einige Anschaffungen.
	A man saves money on some purchases.
	Er bedauert *das Ausgeben* von 14 Euro.
	He regrets ***the spending*** *of 14 euros. (Glossing)*
	He regrets spending 14 euros (Translation)
	Er hatte 61 Euro gehabt. Wie viel Geld hat er am Ende?
	He had 61 Euros. How much money does he have in the end?
C2	Ein Mann spart Geld für einige Anschaffungen.
	A man saves money on some purchases.
	Er hat 18 Euro ausgegeben.
	He spent 18 euros.
	Er hatte 41 Euro gehabt.
	He had 41 euros.
	Wie viel Geld hat er am Ende?
	How much money does he have in the end?
C3	Ein Mann spart Geld für einige Anschaffungen.
	A man saves money on some purchases.
	Er hat jetzt 56 Euro.
	He has 56 euros now.
	Er hat sich über *das Verdienen* von 28 Euro gefreut.
	He was about ***the earning*** *of 28 euros happy. (Glossing)*
	He was happy to earn 28 euros. *(Translation)*
	Wie viel Geld hatte er am Anfang?
	How much money did he have in the beginning?
C4	Ein Mann spart Geld für einige Anschaffungen.
	A man saves money on some purchases.
	Er hat 36 Euro verdient.
	He earned 36 euros.

Er hat jetzt 52 Euro.

He has 52 euros now

Wie viel Geld hatte er am Anfang?

How much money did he have in the beginning?

C5 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er bedauert **das Ausgeben** von 12 Euro.

*He regrets **the spending** of 12 euros. (Glossing)*

He regrets spending 12 euros. (Translation)

Er hatte 64 Euro gehabt.

He had 64 euros.

Wie viel Geld hat er am Ende?

How much money does he have in the end?

C6 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er hat 56 Euro.

He has 56 euros.

Er hat 31 Euro ausgegeben.

He spent 31 euros.

Wie viel Geld hat er am Ende?

How much money does he have in the end?

C7 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er hat jetzt 58 Euro.

He has 58 euros now.

Er hat sich über **das Verdienen** von 37 Euro gefreut.

*He was about **the earning** of 37 euros happy. (Glossing)*

He was happy about earning 37 euros. (Translation)

Wie viel Geld hatte er am Anfang?

How much money did he have in the beginning?

C8 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er hat jetzt 76 Euro.

He has 76 euros now.

Er hatte 13 Euro verdient.

He had earned 13 euros.

Wie viel Geld hatte er am Anfang?

How much money did he have in the beginning?

C9 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er hatte 17 Euro.

He had 17 euros.

Er freut sich über **das Verdienen** von 58 Euro.

*He is happy about **the earning** of 58 euros. (Glossing)*

He is happy about earning 58 euros. (Translation)

Wie viel Geld hat der Mann jetzt?

How much money does the man have now?

C10 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er hatte 34 Euro.

He had 34 euros.

Er verdient 49 Euro.

He earns 49 euros.

Wie viel Geld hat der Mann jetzt?

How much money does the man have now?

C11 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er bedauert **das Ausgeben** von 65 Euro.

*He regrets **the spending** of 65 euros. (glossing)*

He regrets spending 65 euros. (translation)

Er hat jetzt 18 Euro.

He has 18 euros now.

Wie viel Geld hatte der Mann?

How much money did the man have?

C12 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er hat 19 Euro ausgegeben.

He spent 19 euros.

Er hat jetzt 32 Euro.

He has 32 euros now.

Wie viel Geld hatte der Mann?

How much money did the man have?

C13 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er hatte 24 Euro.

He had 24 euros.

Er freut sich über **das Verdienen** von 51 Euro.

He is happy about ***the earning*** *of 51 euros. (glossing)*

He is happy about earning 51 euros. (translation)

Wie viel Geld hat der Mann jetzt?

How much money does the man have now?

C14 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er hatte 82 Euro.

He had 82 euros.

Er verdient 15 Euro.

He earns 15 euros.

Wie viel Geld hat der Mann jetzt?

How much money does the man have now?

C15 Ein Mann spart Geld für einige Anschaffungen.

A man saves money on some purchases.

Er bedauert ***das Ausgeben*** von 52 Euro.

He regrets ***the spending*** *of 52 euros. (glossing)*

He regrets spending 52 euros. (translation)

Er hat jetzt 24 Euro.

He has 24 euros now.

Wie viel Geld hatte der Mann

	How much money did the man have?
C16	Ein Mann spart Geld für einige Anschaffungen.
	A man saves money on some purchases.
	Er hat 74 Euro ausgegeben.
	He spent 74 euros.
	Er hat jetzt 23 Euro.
	He has 23 euros now.
	Wie viel Geld hatte der Mann?
	How much money did the man have?

Note. The sentences shown in italics provide English translations of the German examples used in the experiment. For the sentences with the nominal form, both a word-by-word glossing and a translation is provided to clearly indicate both the linguistic characteristics of the German example and its meaning. C1-C16 corresponds to the factors in Table 2.

The 320 WPs contained ten different templates (see Table 3). Each template was structured in the same format and included four sentences. Each template contained all the factors. The order of problems was systematically varied to avoid ordering effects.

Table 2 - 3 *Templates.*

	Templates
1	Eine Marktfrau verkauft Äpfel auf dem Markt. *A market woman sells apples in the marketplace.*
2	Ein Mädchen hat ihren Freunden Bücher mitgebracht. *A girl brought books to her friends.*
3	Einige Leute wurden zur Party eingeladen. *Some people were invited to the party.*
4	Ein Vater spielt mit Kindern beim Geburtstag Versteckspiel. *A father plays with children at the birthday hide-and-play game.*
5	Eine Studentin muss Wörter lernen. *A student must learn words.*
6	Ein Mann spart Geld für einige Anschaffungen. *A man saves money on some purchases.*

7	Ein Dieb hat einer Frau einige Diamanten gestohlen. *A thief stole some diamonds from a woman.*
8	Eine Tennisspielerin spielt auf der Weltprofitour Turniere im Einzel. *A tennis player plays in the tournaments in single.*
9	Peter möchte ein gebrauchtes Fahrrad kaufen. *Peter would like to buy a second-hand bike.*
10	Ein paar Freunde suchen Pilze im Wald. *A few friends are looking for mushrooms in the forest.*

Each WP consisted of four sentences, which were presented simultaneously (for examples, see Table 2). The second and third sentences contained the two numbers and cue words necessary for the operation. The final fourth sentence contained the question.

The arithmetic operation was limited to addition or subtraction. All problems consisted of two-digit numbers in Arabic notation and overall problem size, as the number of various strategies increases with problem size (Verschaffel, De Corte, Lamote, & Dherdt, 1998), was matched between non-carry and carry addition, as well as between subtraction and addition, problems (problem size: Subtraction M = 76.7, SD = 15.1; Addition: M = 82.5, SD = 18.8). Number pairs for the tasks were matched in difficulty because calculations with smaller numbers have been shown to be retrieved from a network of mental representations, while calculations with larger numbers requires a transformation process (LeFevre, Sadesky, & Bisanz, 1996). Therefore, because it has been shown that response times on simple arithmetic problems are in general slower and more error prone if the operands and their correct solutions are larger, we have controlled the problem size in our experiments (Klein et al., 2009).

Number pairs with identical numbers, "shot numbers", with a zero in the units digit, mirror numbers (e.g., 24 - 42), and all combinations of the operand with the same numbers in the tens or units digits (e.g., tens: 41 + 43, units: 24 - 14) were excluded to prevent any automatic mental retrieval. The order in which the number size was presented in the task was balanced across all passes in equal proportions: Half of the stimuli started with the larger number and the other half with the smaller number, because although 4 + 2 = 6 and 2 + 4 = 6 are arithmetically equivalent, the processing may differ (Kaput, 1979). For instance, Nys (2010) found that test subjects preferred when the larger number occurred first in mental addition. The number of tasks with even or odd results was balanced as well because parity also influences the difficulty of simple addition and subtraction (Hines, 2013).

Lexical consistency or inconsistency was encoded through an operation lexically evoked in the text. The second or the third sentence contained a cue word (in verbal or nominal form), which evoked an operation (e.g., “sell”/“selling” for subtraction). Depending on the scenario of the story, the cue word could be misleading. For example, the German verb “leihen” can mean "to lend" or "to borrow", which from the point of view of the lender evokes subtraction but from that of the borrower evokes addition. To make the perspective explicit, a reference introductory sentence was included to provide context before every task.

The type of nominalization in the text was formulated in the so-called infinitive-based nominalization. Generally, in nominalization the words lose their verbal characteristics and behave like real nouns (Hamm & van Lambalgen, 2002). In German, there are two possibilities for creating nominal forms: one using the “–ung” suffix (e.g., “landen” “to land” becomes “Landung” “landing”) and the other based on the infinitive (“landen” becomes “das Landen”). The latter is very close in meaning to the underlying verbs, denoting the events or states that the verbs denote (Scheffler, 2005). They can be formed for any German verb, whereas “–ung” nominalization is not available to all verbs. For preparing the stimuli, it was important that the nominalized form be applicable to all selected keywords (in our case verbs), and that the meaning stayed as close to the original meaning as possible, so we systematically used infinitive-based nominalization instead of “–ung” nominalization. All sentences were constructed with active verb forms given that active voice has been shown to facilitate WP understanding (Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005), and we wanted to avoid additional linguistic complexity. Compared to other studies (e.g., Prediger, Wilhelm, Büchter, Gürsoy, & Benholz, 2015), which changed several noun phrases at once in the text, we kept the number of changed noun phrases as small as possible. For example Spanos et al. (1988) showed that certain grammatical features (e.g., prepositional phrases, noun phrases) prevented participants from fully understanding mathematical WPs. In our study, the nominalized and verbal forms differ only as much as is necessary to realize the variants: The number of noun phrases differ because in one case a verb is used (e.g., verkauft), whereas in the other that verb is nominalized (e.g. den Verkauf) and embedded under another verb. Therefore, the sentence with the nominal form is syntactically more complex. Other conditions do not differ in syntactic complexity.

To minimize the risk that response time would be adversely affected by prolonged reading and the extra burden of working memory (Shaftel et al., 2006), filler words and information not relevant to understanding (Muth, 1992) and solving the task were avoided. We kept the length, word count, and frequency as similar as possible given that lexical complexity (word frequency) and syntactic complexity (as reflected by the mean sentence length in words, item length in words, noun phrase length, and number of prepositional phrases (Abedi et al., 2005)) influence reading. The overall number of words and average sentence length are related to reading difficulty with longer sentences posing greater challenges for readers (Butler, Bailey, Stevens, Huang, & Lord, 2004). Item length shows relatively consistent negative effects, increasing item difficulty (Abedi et al., 1997). Table 4 provides the average word count, character length, syllable count, character count per word, and syllable count per word for each condition as well as frequency measures. Frequencies were calculated using the COW German Corpus (Schäfer, 2015; Schäfer & Bildhauer, 2012).

Table 2 - 4 *Linguistics Measures Across Conditions. Word Count, Total Character Length, Total Syllable Count, F_logpermil_10 (the log10 of the frequency per million shifted into the positive range by adding 10), Band (frequency band).*

	Word Count		Total Character Lenght		Total Syllable Count		F_logpermil_10		Band	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Addition										
Non Carry/Non-Borrow										
Consistent Form										
Verbal Form	26.15	3.47	127.85	24.14	41.15	8.74	128.14	10.58	7.85	0.53
Nominal Form	30.00	3.42	145.55	22.91	46.90	8.19	130.9	11.54	7.70	0.53
Inconsistent Form										
Verbal Form	25.15	2.60	128.9	20.77	40.85	7.01	123.42	8.72	8.13	0.44
Nominal Form	28.5	3.14	145.4	22.18	45.65	7.56	129.62	7.30	7.99	0.25
Carry/Borrow										
Consistent Form										
Verbal Form	26.15	3.47	127.85	24.14	41.20	8.70	128.27	10.57	7.84	0.53
Nominal Form	30	3.42	145.55	22.91	46.90	8.19	130.90	11.54	7.70	0.53
Inconsistent Form										
Verbal Form	25.15	2.60	128.90	20.77	40.85	7.01	123.42	8.72	8.13	0.44
Nominal Form	28.45	3.15	145.40	22.18	45.65	7.56	129.62	7.30	7.99	0.25
Subtraction										
Non Carry/Non-Borrow										
Consistent Form										

Verbal Form	25.45	1.57	128.95	15.37	40.25	4.18	124.77	10.43	8.10	0.46
Nominal Form	29.15	2.87	146.35	18.05	45.50	5.76	131.74	7.86	7.95	0.28
Inconsistent Form										
Verbal Form	26.45	2.78	130.90	19.74	41.70	7.31	127.36	10.73	7.90	0.58
Nominal Form	30.25	3.13	147.30	18.67	46.75	6.45	130.55	11.04	7.77	0.54
Carry/Borrow										
Consistent Form										
Verbal Form	25.45	1.57	128.95	15.37	40.25	4.18	124.77	10.43	8.10	0.46
Nominal Form	29.15	2.87	146.35	18.05	45.50	5.76	131.74	7.86	7.95	0.28
Inconsistent Form										
Verbal Form	26.45	2.78	130.90	19.74	41.70	7.31	127.36	10.73	7.90	0.58
Nominal Form	30.3	3.08	147.45	18.45	46.80	6.38	130.73	10.89	7.76	0.53

Additionally, reading speed, reading comprehension, calculation for subtraction and addition, verbal working memory, visual working memory and central executive were assessed for every individual. Because of length restrictions and because the N was too low in this study to derive strong conclusions about individual differences from multiple correlations these results are only reported in the Supplementary Material 1. for the interested reader, but not discussed in the article itself. Generally, the correlations were in the expected direction; i.e., better reading and calculations skills were related to better performance on the WPs.

PROCEDURE

The experiment took place in a laboratory at the University of Tübingen. The stimulus was presented on a 19-inch monitor in white Times New Roman font, size 24, bold on a black background. Voice or spoken response was used to detect response time. For this purpose, 32 two-digit numbers were presented on the screen, divided into two blocks. For this, the Voice Key (Creative EMU 0202, USB audio interface) was used as a trigger, consisting of the interface and a headset with microphone. To reduce potential measurement inaccuracy (Kessler, Treiman, & Mullennix, 2002) or at least keep it constant, all participants were asked to loudly and clearly say, "is," followed by the displayed number (e.g., in German "ist 45"). Subjects were instructed that only mental calculations were allowed and that no other calculation support (such as finger movements) should be used.

The 320 tasks were divided into two blocks, 160 tasks each, with a break in-between. Each task was presented separately in pseudo-randomized order. Presentation was adjusted so that in each block all 16 conditions were included. In addition, they were mixed, and a single condition could be presented no more than three times in a row. Participants in the experiment were seated such that the distance between their eyes and the monitor, on which the WPs were displayed, was approximately 70 cm. All four sentences were presented at the same time on the screen, aligned in the centered position, with one line containing one sentence. The stimuli were present until the Voice Key was triggered by the "Is(t)” cue. The response time was taken from the point that the stimuli was presented till the Voice Key was triggered. After the trigger, the participants saw a black background and communicated the solution. The investigator noted the response and other incidents during the experiment. A fixation point before each trial was used (x/y coordinates: 112/384).

RESULTS

ANALYSIS

Statistical analysis of RTs and ACC data was performed using the R-project statistical computing software (Team, 2014). Response times were analysed using linear mixed effects models (LMM), and ACC data were analysed as a binomial variable using generalized linear mixed effects models (GLMM) as implemented by the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). In all of the subsequently considered models independent random intercepts are included for both participants and items. Random slope models turned out to be too complex for the current data set, because estimation algorithms did not converge. Each factor was dummy coded, with the easier condition (forming the reference category) labelled by 0, and the more difficult condition by 1 (subtraction operation, inconsistent lexical form, carry/borrow, nominalized form). This means that the intercept corresponds to the mean response time (mean ACC, respectively) of an item, which is in the reference category for all of the factors. Models were built up stepwise in a hierarchical way, and statistical decisions are based on incremental likelihood ratio tests (LR Chisq in subsequent tables). Beyond the random effects, the null model (Model 0) included an intercept only, and no other fixed effects. Model 3 formed the full model including all interactions for both RTs and ACC. Table 5 provides means and standard deviations for all factor level combinations for both response time and ACC.

Table 2 - 5 *Response Time, Accuracy Mean and Standard Deviation for all 16 Factors.*

	Reaction Time		Accuracy in %	
	M	SD	M	SD
Total	9.37	2.54	0.93	0.03
Addition				

Non Carry/Non-Borrow						
	Consistent Form					
		Verbal Form	8.31	2.22	0.95	0.08
		Nominal Form	8.76	2.79	0.95	0.05
	Inconsistent Form					
		Verbal Form	8.57	2.19	0.92	0.06
		Nominal Form	8.81	2.35	0.94	0.07
Carry/Borrow						
	Consistent Form					
		Verbal Form	9.26	2.21	0.95	0.07
		Nominal Form	9.94	2.57	0.93	0.06
	Inconsistent Form					
		Verbal Form	9.55	2.77	0.92	0.07
		Nominal Form	9.66	3.00	0.89	0.10
Subtraction						
Non Carry/Non-Borrow						
	Consistent Form					
		Verbal Form	8.67	2.53	0.95	0.06
		Nominal Form	8.69	2.35	0.95	0.06
	Inconsistent Form					
		Verbal Form	9.09	2.69	0.93	0.05
		Nominal Form	9.55	2.51	0.92	0.08
Carry/Borrow						
	Consistent Form					
		Verbal Form	10.07	3.12	0.88	0.10
		Nominal Form	10.10	3.21	0.93	0.05
	Inconsistent Form					
		Verbal Form	10.46	2.55	0.91	0.06
		Nominal Form	10.76	3.39	0.92	0.08

RESPONSE TIME

Adding the main effects carry, operation, lexical consistency, and nominalization to Model 0 one by one showed that all of them improve model fit. Model 1 included all of these main effects, and formed the basis for testing interactions, which again were tested in a stepwise manner. Model comparison (see Table 6) revealed that the most parsimonious model exhibiting a good fit to the data was Model 2, containing all main effects as well as the interaction operation * lexical consistency. Including any additional interactions did not improve fit.

Table 2 - 6 *Comparison of the LMMs in Response Time Analysis.*

	Effects	Df	AIC	BIC	LR Chisq	Δ df	p-value
Model 0	intercept only	4	32753	32781			
Model 1	main effects added	8	32644	32698	117.48	4	< 2.2e-16
Model 2	Operation*Consist. interaction added	9	32640	32702	5.52	1	0.019
Model 3	full model	19	32654	32783	6.60	10	0.763

The estimates for Model 2 are listed in Table 8, together with standard errors and t-values. Consistent with our first hypothesis, the following main effects turned out to be significant (see Figure 1). Carry produced the most pronounced effect, increasing latencies by 1.16 sec compared to problems without carry. Subtractions were only slightly slower than additions on average (0.30 sec), and nominalization provided about the same effect size (0.28 sec).

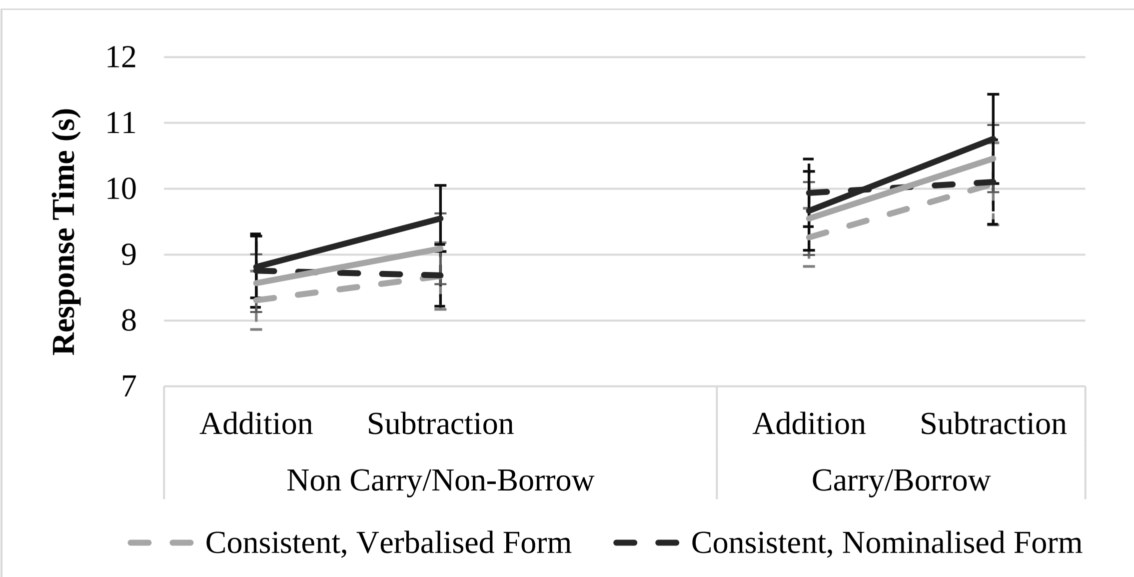

Figure 2 - 1 Response Times Separated for each Factor. Error bars show standard error of the mean (SEM). Dark line is addition, grey is subtraction; solid line: carry/borrow, dashed line: non-carry/non-borrow factors.

The two-way interaction (see Figure 2) between the arithmetic factors operation and lexical consistency (0.54 sec) indicated that there was a consistency effect (difference between lexically inconsistent form vs. lexically consistent form) for subtraction problems only (RTs increase by 0.60 sec on average). For addition problems, lexical consistency had no effect (0.06 sec). Random effects indicate that there was substantial individual variation in mean RT across participants (SD=2.38), but items seemed to be quite homogenous (SD=0.84).

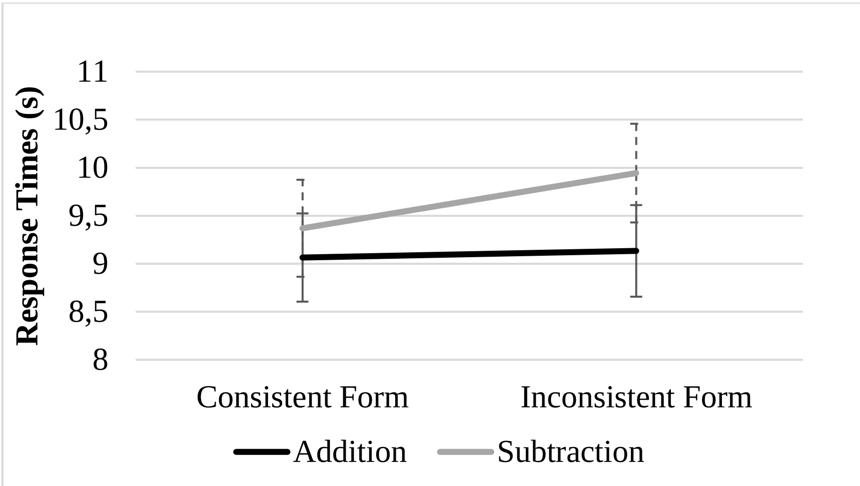

Figure 2 - 2 Two-way interaction between factors Operation * Lexical Consistency. Error bars show standard error of the mean (SEM).

In sum, the current results suggest that all factors wielded influence on the latencies, which for carry, operation, and nominalization can be described as main effects. Consistency effects were present for subtractions only.

ACCURACY

Comparing GLMM models in the same way as the LMM models for RTs reveals (see Table 7) that the model that describes the data best is Model 1. Its estimates are collected in Table 8. Model 1 includes fixed main effects for carry (log odds -0.36) and lexical consistency (log odds -0.29) only, both are negative and thus are decreasing ACC. Adding the other main effects (Model 2), or any interaction (Model 3) did not improve model fit. However, care should be taken in interpreting the ACC results, due to the overall high ACC (92%) indicating a ceiling effect (log odds 3.14 for the intercept). This is also reflected by the comparatively small individual variation captured by the random effects of participants (SD= 0.36) and items (SD= 0.65).

Table 2 - 7 *Comparison of the GLMM Models for Accuracy.*

	Effects	df	AIC	BIC	LR Chisq	Δ df	p-value
Model 0	intercept only	3	3753.1	3773.8			

Model 1	Carry & Consistency main effects added	5	3742.9	3777.4	14.24	2	<0.001
Model 2	all main effects added	7	3746.3	3794.7	0.60	2	0.740
Model 3	full model	18	3755.1	3879.5	13.18	11	0.282

Table 8 presents the model outputs for the final models for RT (Model 2) and Accuracy (Model 1).

Table 2 - 8 *Model Outputs for Response Time (Model 2) and Accuracy (Model 1).*

	Response Time (sec), Model 2			Accuracy (Log Odds), Model 1		
Random Effects	Variance	SD		Variance	SD	
Item	0.71	0.84		0.42	0.65	
Subject	5.67	2.38		0.13	0.36	
Residual	6.69	2.59		–	–	
Fixed Effects	Estimate	SE	t	Estimate	SE	Z
Intercept	8.38	0.50	16,90	3.14	0.13	23.24
Operation	0.30	0.16	1,86	–	–	–
Lexical Consistency	0.06	0.16	0,36	-0.29	0.12	-2.44
Carry/Borrow	1.16	0.11	10,27	-0.36	0.12	-3.05
Nominalization	0.28	0,11	2,50	–	–	–
Operation * Lexical Consistency	0.54	0,23	2,36	–	–	–

DISCUSSION

The main goal of this study was to explore key linguistic and arithmetic factors that may contribute to the difficulty many people experience when solving WPs. We manipulated linguistic complexity to be independent and orthogonal to the arithmetic complexity of the underlying WP in such a way that there was one factor related to arithmetic/linguistic complexity and one that was unrelated. For example, the linguistic factor lexical consistency and the arithmetic factor operation were related to arithmetic complexity, whereas the linguistic factor nominalization and the arithmetic factor carry/borrowing were not. We examined the main effects and interactions of these manipulations on WP performance and their connection with the problem-solving process.

In our first hypothesis, we expected that all linguistic and arithmetic factors would have a main effect on performance (i.e., response time). The factors carry, operation and nominalization all had a main effect on response time. Confirming the H1.1 hypothesis, it took participants more time to solve subtraction tasks as compared to addition tasks. Additionally, the main effect of carry indicated that participants had a greater response time when solving WPs with a carry/borrow operation as compared to tasks without this operation, which is in line with the H1.2 hypothesis. This is consistent with and extends similar findings in which arithmetic calculations were not embedded in a WP (Geary et al., 2012; Imbo et al., 2007). This was also expected because the carry effect in addition tasks and the borrow effect in two-digit subtraction tasks are both associated with increased arithmetic difficulty (Deschuyteneer et al., 2005). The H1.3 hypothesis, that WPs with an inconsistent form should take significantly longer to answer than those that are consistent, was partially confirmed, because the consistency effect was present only in the WPs with subtraction. Here, our findings are consistent with the results of Hegarty et al. (1995) and Van der Schoot et al. (2009); and opposed to those of Verschaffel et al. (1992), who found a consistency effect in children but not in adults. Finally, the nominal form significantly increased response time compared to the verbal form, which confirms the H1.4 hypothesis. This result extends previous findings that nominalizations can be a source of difficulty (Abedi et al., 2005; Prediger et al., 2015). Contrary to studies that concluded that difficulty arises only in complex multi-step WPs with a high density of nominalization (Schlager et al., 2017) our study shows that simple nominalization influences the complexity of the WPs as well. Considering that the number of sentences, words, and even characters in this study were kept consistent between WPs, this finding provides support for the importance of unrelated linguistic factors. Generally speaking, the main effects support the direct influence of stimulus attributes on WP performance as in the theoretical process model proposed by Daroczy et al. (2015).

Interaction of Linguistic and Arithmetic Factors

In the same model (Daroczy et al., 2015), however, we hypothesized that WP difficulty comprises not only the linguistic complexity of the text and arithmetic complexity of the arithmetic problems but also the interaction of these factors because some attributes are processed at common stages. Therefore, we expected the interactions between linguistic and arithmetic factors to be more pronounced for the related factors than for the unrelated factors. As expected in the H2.1 hypothesis, operation interacted with lexical consistency. The interactions were particularly pronounced for factors relating to both text and arithmetic operation. There was a consistency effect for subtraction but not in the case of addition, suggesting an over-additive effect in the most difficult condition. The direction of this interaction between operation and lexical consistency is supported by some studies but not by others. For example, in this study the direction of the interaction was the opposite to that previously found by Verschaffel et al. (1992), where the difficulty of overcoming inconsistent language was enhanced in the case of addition and multiplication. An explanation as to why our results are different could be that the previous study used compare problems, which have a different semantic background from the change problems used in the current study. Although the interaction between lexical consistency and operation is quite stable across various studies, in ACC there was no interaction of these factors, it is important to note that the overall error-rate was very low, which might indicate a ceiling effect. A possible explanation for this over-additive interaction is given by the model of Daroczy et al. (2015). Joint domain-general stages of processing like working memory have limited resources. Difficulty in both the linguistic and the arithmetic domain leads to particularly slow processing, perhaps because there may not be enough resources. For this interaction between operation and lexical consistency, however, another alternative interpretation is possible, using linguistic markedness. An interaction between markedness and lexical consistency in these studies is reflected in the fact that problem solvers find it especially difficult to solve a problem in the inconsistent-marked condition, i.e., inconsistent addition problems seem to be harder than inconsistent subtraction problems (Hegarty et al., 1995; Verschaffel et al., 1992). This interaction between operation and lexical consistency is explained by the fact that increased difficulty does not depend on the operation but on the greater semantic complexity of the marked sentences. Nevertheless, the problem with this interpretation is – unlike the factor nominalization in this study – markedness cannot be manipulated independently from lexical consistency and operation (Van der Schoot et al., 2009). Namely, consistent and marked-inconsistent problems always concern addition, whereas the two other problem types (unmarked-inconsistent and marked-consistent) always concern subtraction, so it is still unclear whether the semantic change or the interaction of operation and language causes this difficulty. This needs to be disentangled in the future. Our findings suggest that the interaction between operation and lexical consistency does not only depend on the semantic and arithmetic features, but also the interaction of language and mathematics.

Although we have not found a significant interaction between nominalization and other factors in the linear mixed effect models, in the case of nominalization, the consistency effect for response times was slightly more pronounced for subtraction problems than for addition problems, something we did not observe in the verbal forms. Therefore, we suppose that in the linguistically less demanding conditions there are enough cognitive resources to process the complex lexical consistency and the complex operation condition in combination. However, when the text gets linguistically difficult (which requires additional resources), there are no resources left for this most demanding interaction condition, which leads to slower responses. This might especially hold for individuals with lower cognitive abilities or who are in an earlier developmental stage, children, for example.

Finally, in the case of carry, i.e., for the arithmetic factor, which was hypothesised to interact less with linguistic factors, because it was unrelated, there was no interaction with other linguistic factors. This partially supports the hypothesis that for factors affecting only arithmetic the interactions may be absent or less consistent (H2.2).

In sum, the interaction between linguistic and arithmetic factors speaks to possible joint sources of linguistic and arithmetic difficulty. Carry produced the most pronounced effect; subtractions were only slightly slower than additions on average and nominalization had about the same effect size on the response time compared to operation. The consistency effect was significant for subtraction problems only.

UNDERLYING COGNITIVE PROCESSES IN SOLVING ARITHMETIC PROBLEMS

Problem-solving Models. The proposed models for how WPs are solved disagree on the origin of the internal problem model, i.e., whether the internal problem representation is a result of a schema or a situation – i.e., mental representation model. The findings of this study support the situation or mental representation model over the schema model (Thevenot & Barrouillet, 2015; Thevenot & Oakhill, 2005). Provided that in the schema model only the keyword matters and the difficulty of the inconsistent WP results from the mismatch between text and schema (Kintsch & Greeno, 1985), it may be suggested that in this study only nominalization and lexical consistency should result in significant differences (when only the keyword is affected) if this was the appropriate model. On the contrary, we found effects for all other factors, as well as an interaction. Additionally, the WPs with nominalized, inconsistent form and subtraction resulted in the highest response time, which can be interpreted as the additional text difficulty making it harder to create a problem model in this mathematically more difficult condition. In fact, the interaction also holds for the situation model that requires the construction of a mental representation of the situation described by the problem (Johnson-Laird, 1983).

Interactions between Related and Unrelated Factors and the Problem-solving Models. Additionally, the problem-solving models mentioned above did not agree on whether the problem-solving phases are fully separable or not. Manipulating related and unrelated factors at the same time might provide elaboration on how the presence or absence of an interaction between the factors could support the existing models mentioned above. In particular whether or not the initial reading phase, and the last calculation phase, interact with other problem-solving phases, like the mental representation for example, could be addressed in future studies.

The relation of calculation phase to other phases of problem-solving: In this study we have manipulated computation with two factors. In one the number difficulty (i.e., carry) was changed, and in the other the difficulty of the operation was changed to gain a better understanding of whether the computational process affects the mental model or not. If computation also affects the problem model phase, where the mental model is built, we would have found an interaction between the unrelated arithmetic factor carry and the other factors. However, we found no interactions between carry and the other linguistic and arithmetic factors. These results are favored by propositional theory, which sees the calculation phase and the other problem-solving phases as distinct. This finding is in line with Rabinowitz and Wooley (1995) who found no interaction for either response time or for ACC between the factors problem size, carry and others. This means that in this study we found no evidence that number difficulty would affect another stage of problem-solving – i.e., the mental representation, besides the calculation phase. Therefore, in the case of carry, sequential processing is highly probable, and it likely affects the problem-solving phase after the creation of the mental model and does not influence the quality of the representation. However, there might be an alternative interpretation. Failing to solve a WP successfully is priory hypothesized due to the failing of the creation of the mental model (Hegarty et al., 1995; Verschaffel et al., 1992). Nevertheless, the factor carry affected the correctness of the solution. This means that it might be hard to determine from a non-correct solution if it is due to the 1) the correct mental representation and wrong calculation or 2) to the incorrect mental representation because, for example, the incorrect calculation can be primed by specific words (Bassok, Pedigo, & Oskarsson, 2008). A second alternative explanation would be that the carry operation also gets more difficult as textual processing gets more difficult but the missing interaction with other factors are at odds with this explanation.

The relation of reading phase to other phases of problem-solving: The arithmetically unrelated linguistic factor (i.e., nominalization) that did not change the underlying mathematics did not show an interaction with other factors. However, an interaction between nominalization and the other factors would have suggested that the reading comprehension phrase interacts with other problem-solving phases. The results again favor the propositional model over the non-sequential cyclic models. This is not exactly in line with the hypothesis derived from the model of Daroczy et al. (2015) which would expect an interaction also in case of factors where the cognitive load increases. The missing interaction between unrelated linguistic and arithmetic factors could also mean that it is not correct to assume a limited load domain-general stage model as described in the model of Daroczy et al. (2015). Nevertheless, nominalization showed a main effect on response time that confirms the direct influence route from the model. Therefore, a possible interpretation is that the arithmetically unrelated linguistic factor (i.e., nominalization) might affect only the initial reading phase. The ACC results support this, as there was no significant main effect for nominalization, nor an interaction with other factors. On the other hand, the interaction between lexical consistency and operation might mean that in the case of lexical consistency the process of reading is not completely separable from the process of solving. However, this interaction between lexical consistency and operation is consistent with both the sequential and cyclic models.

In summary, we can say that the absence of an interaction between the unrelated and related factors is more consistent with the propositional theory than with the cyclic model. This might imply that the initial reading phase, the calculation phase and the building of the problem model phase can be viewed as distinct, non-overlapping stages of problem-solving. As no interaction was found between carry and nominalization or between other factors, these factors might influence different processing stages – cf., Sternberg (1969). Number difficulty (i.e., carry) seemed to affect the calculation/execution phase, and text difficulty (i.e., nominalization) the initial text comprehension phase.

THEORETICAL IMPLICATIONS

Mathematical texts have their own terminology and language, and it is claimed that they require special literacy skills that are developed across years (Burton & Morgan, 2000). Additionally, it is suggested that WPs and calculations represent distinct domains of mathematical performance

(Fuchs et al., 2008; Swanson, 2004). Language is assumed to have a stronger effect on WP solution success than, for example, calculation skills (e.g., Fuchs et al., 2014). This hypothesis is supported by the documented strong connection between text comprehension and WP-solving (Boonen et al., 2013; Boonen et al., 2014; Kintsch & Greeno, 1985; Vilenius-Tuohimaa, Aunola, & Nurmi, 2008) and by the observation that some students who have otherwise little to no issues with arithmetic tasks cannot solve those that are written in textual form (Nesher & Teubal, 1975). Because of this evidence, WP research and intervention often focus mainly on text difficulty and the problem of constructing an adequate representation (Cummins et al., 1988; Davis-Dorsey et al., 1991; De Corte et al., 1985; Lewis & Mayer, 1987). Therefore, recent studies suggest a stronger focus should be placed on training mathematical literacy and reading skills (e.g., Fuchs et al., 2008). In our study, we observed interactions between arithmetic factors and linguistic factors but also that arithmetic difficulty plays a strong role on its own. Therefore, we suggest it might be an incomplete approach to mainly focus on text difficulty in WPs. Instead, mathematical and language requirements need to be considered together when designing WPs. Therefore, in addition to better text comprehension instruction when teaching WP-solving strategies, we suggest that the involvement of arithmetic factors should not be neglected, even if children are able to solve equivalent mathematical calculations in less resource-demanding tasks. Instruction in text comprehension or mathematics may even need to be individually adapted. In some types of WPs, arithmetic difficulty might play a role, while in others linguistic difficulty is more important. Our suggestion is to confront children with a particular type of WP, namely, the problem from which they can learn most.

LIMITATIONS OF THE STUDY AND FURTHER PERSPECTIVES

Integral to this study is the systematic investigation of select linguistic and arithmetic factors to distinguish from where the difficulty of WPs originates. For example, when faced with two WPs with the same arithmetic operations but different textual difficulties an individual may not be able to create a problem representation for the textually more difficult problem. Another outcome might be that despite the correct problem representation and otherwise good mathematic skills a calculation error occurs due to the increased cognitive load from the interaction between linguistic and arithmetic factors. In both cases the result is the same, namely an incorrect solution. However, each case would require a different intervention. Therefore, we suggest that further elements of text characteristics that cause difficulty for various individuals should be systematically investigated (See Supplementary Material). Other linguistic and arithmetic factors that are potential candidates for systematic study include number sense (Dehaene, 2001), the presented order of the numbers, and the influence of these features on solution ACC, paired with cognitive ability and item difficulty. For instance, this study did not include factors that did not change the underlying structure but manipulated the calculation process. For example, Bassok et al. (1998) have shown that calculations represented by functionally connected words are easier to add up. Such a linguistic factor could be a candidate to show an interaction between the calculation phase and the reading phase. This is supported by (Bagnoud, Burra, Castel, Oakhill, & Thevenot, 2018) who found that brain activity differs when exposed to discrete quantities or continuous quantities (e.g., apples, meter, rope).

It is important to note that the current paper did not cover and investigate all the elements of the proposed theoretical WP-solving process model from Daroczy et al. (2015) because the model also suggests a connection not only to task characteristics but to individual abilities or to individual strategies, as well as the inclusion of the role of the environment. The current study was conducted in adults to examine the factors that influence WP difficulty. However, selecting participants from other groups might lead to different results. For instance, examining a group of children, who, when compared to adults, have a more limited working memory capacity and therefore experience higher cognitive load (Swanson & Beebe-Frankenberger, 2004) might lead to more over additivity than that observed in adults. Finally, it is important to note that the WP is a didactical construct which serves purposes in mathematical education from simple exercises on basic operations (Greer, 1997) to more complex tasks (Verschaffel et al., 1997). Performance when solving WPs is highly influenced by the environment and what is expected in the learning scenario (Cobb & Bauersfeld, 1995). Therefore, in future diagnostic assessment both the item characteristics (i.e., the linguistic and arithmetic complexity) and the individual characteristics (i.e., general linguistic and mathematical skills), as well as the environment, should be considered.

Conclusion

In the current study, we investigated the influence of linguistic and arithmetic factors on WP performance. We argue that it is essential to investigate the influence of linguistic and arithmetic factors on WP performance because: (1) the connection between mathematical and linguistic factors gives rise to the difficulty of WPs, and (2) one needs to distinguish factors where linguistic and mathematical aspects are conceptually linked (e.g., consistency of verb meaning with the mathematical operation to be performed) from linguistic and arithmetic factors that are not linked in the other domain (e.g., nominalization, carry) to identify if or when interaction effects are to be expected. Our results indicate that both linguistic and arithmetic complexity contribute to the difficulty of a WP and that linguistic and arithmetic factors interact. Therefore, linguistic and arithmetic complexities are not always fully separable attributes.

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments.

Compliance with Ethical Standards

Informed consent was given by all participants. The study was performed in accordance with the ethical standards of the Declaration of Helsinki.

Disclosure of potential conflicts of interest: The authors declare that there is no conflict of interest.

SUPPLEMENTARY MATERIAL: CHARACTERISTICS OF THE INDIVIDUAL

In the main article, we have been concerned with the characteristics of WPs, i.e. their linguistic and arithmetic complexity. However, individual cognitive abilities such as reading skills, working memory, mathematical ability and other skills play an important role in WP success. It is still an open question which above mentioned components play a role in different types of WPs. For instance, linguistic capabilities could be more important for linguistically complex WPs, while arithmetic capabilities could be more important for arithmetically complex WPs. Due to the relatively low number of participants, our study is not conclusive here, but it already gives some hints about individual differences, which might be important in WPs. Therefore, we would like to give this information to the interested reader but will not draw strong claims from these data.

The following individual characteristics were studied for the following reasons:

Working memory. It has been suggested that WP performance in general is strongly related to working memory and a strong relationship has been documented between individual differences in WM and performance in arithmetic WPs (Adams & Hitch, 1997; Passolunghi & Siegel, 2001; Swanson, 2004; Swanson et al., 1993). WP-solving is hypothesized to show a stronger correlation with working memory than basic number knowledge (Peng, Namkung, Barnes, & Sun, 2016). Although mathematical skills significantly affect the relationship between working memory and mathematics and for example Imbo et al. (2007) emphasized the role of working memory in mental arithmetic tasks, a number of studies have also shown that the correlation between working memory and problem-solving ACC is substantially lower when corrected for differences in reading comprehension (Fuchs et al., 2006).

Reading and mathematical skills. Successful WP-solving may also be related in part to basic reading and mathematical skills instead of any individual differences in working memory components (Swanson, Lussier, & Orosco, 2015; Zheng et al., 2011). WPs are introduced in schools after children have learned the formal operations of addition and subtraction (De Corte & Verschaffel, 1987), so many papers argue for the importance of reading skills in WP-solving as opposed to arithmetic problem-solving skills. For instance, Fuchs et al. (2006) argue for reading skills alone, as mediator variables between working memory and WP-solving (but see Passolunghi and Siegel (2001) for opposite suggestions).

The mechanisms and implications of both skills are still under debate. On the one hand, reading comprehension is hypothesized to have an indirect effect on WP-solving performance via its influence on relational processing (Lee et al., 2009). On the other hand, studies by Tolar et al. (2012) and Swanson (2006) indicate that processing speed plays a unique role in arithmetic, whereas language comprehension uniquely predicts ACC of WP solutions. Therefore, participants with better reading skills should be less affected by linguistic complexities. Although reading ability is positively related to scores on tests of arithmetic problem-solving (Aiken Jr, 1971), children who have difficulties reading do not always perform poorly in mathematics. At the same time, arithmetic abilities are not often referred to or measured (Boonen et al., 2013). A task that is more demanding in arithmetic processing is also more demanding in the whole solution process of the WP. Therefore, participants with better mathematic skills should be less affected by arithmetic complexity.

Assessments

To examine whether characteristics of the individual, and not only characteristics of the problems, influence WP-solving we assessed reading comprehension and reading speed, mathematical abilities (Fuchs et al., 2015), and working memory. Our hypothesis is that subjects with higher working memory scores, better arithmetic abilities and better language comprehension skills will solve problems faster. We also suggest that individuals with better arithmetic abilities are less affected by arithmetic item complexity and individuals with better linguistic skills are less affected by linguistic item complexity.

Working memory assessment. Two tests provided working memory measures. Verbal working memory performance was tested with the letter span back and forward tests. To test the spatial memory span, a Corsi block-tapping test (Corsi, 1973) was used. After three consecutive errors the test was aborted. See Table 1 for the descriptive measures.

Reading skill assessment. To measure reading abilities, we used a *speed of reading- and reading comprehension test,* which was designed for grades six through 12 (Schneider, Schlagmüller, & Ennemoser, 2007). This test is a power test (so it was considered adequate to measure adults) and measures the speed of reading as well as reading comprehension.

Mathematical skill assessment. To measure basic mathematical skills, we used two short, self-made speed calculation tests for addition and subtraction (Huber, Fischer, Moeller, & Nuerk, 2013). Both tests contained 28 math problems and no text. All problems consisted of two-digit numbers (the pairs of numbers were different from those used for the WPs). Participants had 90 seconds to solve each sheet. As an independent variable, we used the number of correctly solved items.

Table 2S - 1 *Descriptive Measures (N=25).*

		Memory					Mathematical and Reading Skill			
		Verbal Memory		Visual Memory		Central Executive	Mathematical Skill		Reading Measures	
	Age	Letter_Span Forward	Letter_Span Backward	Corsi_Block Forward	Corsi_Block Backward	Central Executive	Addition	Subtraction	Reading Comprehension	Reading Speed
Mean	22..08	8.92	8	5.92	5.84	6.92	21.32	16.8	23.04	981
SD	2.59	1.92	1.77	1.09	0.97	1.05	4.55	3.71	9.2	329.22

Results on the influence of individual capabilities on WP performance

To examine other mechanisms that may have affected WP-solving, correlations of the individual cognitive abilities with the overall mean of the response time were calculated (see Table 2 for a table of the raw correlations between the included variables). The following individual cognitive abilities were included in the analysis: speed calculation tests in addition and subtraction (i.e., the number of correctly solved tasks in 90 s), reading comprehension (points achieved, resulting from the number of correctly identified words), reading speed (number of words read in 4 minutes), working memory (verbal: letters forward, spatial: cube episode forward, central executive: mean value of letter span and Corsi block sequence backwards). We conducted several multiple linear stepwise regressions: Response latency can be negatively predicted from reading comprehension and speed calculation (addition) [R^2 = . 775, adjusted R^2 = .754, $F(2, 22) = 37.87$, p <.000]. The correlation yielded better reading comprehension and mathematical skills result in lower reaction times.

Table 2S - 2 *Pearson Product- Moment Correlations of the Cognitive Variables and Response Latency (N=25).*

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Response Latency	-													
2. Consistency Effect	-.19	-												
3. Nominalisation Effect	-.59**	.28	-											
4. Operation Effect	-.45*	-.01	.52**	-										
5. Carry Effect	-,47*	.06	.45*	.36	-									
6 Reading Comprehension	-.81**	-.35	.56**	.19	.50*	-								
7 Reading Speed	-.77**	.20	.51**	.20	.50*	.96**	-							

8 Corsi Block Forwards	.15	-.15	-.52**	-.16	-.03	-.22	-.17	-						
9 Corsi Block Backwards	-.09	-.02	-.06	.19	,14	-.10	-.15	.42*	-					
10 Letter Span Forward	-.01	-.17	.11	.09	,01	-.00	.01	.035	.20	-				
11 Letter Span Backward	-.26	.11	.31	.14	-.00	.35	.33	-.22	-.04	,40*	-			
12 Central Executive	-.28	.09	.25	.22	.065	.26	.23	.00	.44*	.45*	.88**	-		
13 Speed Test Addition	-.71**	-.03	.29	.52**	.28	.50*	.46*	.08	.33	.04	-.06	.10		
14 Speed Test Subtraction	-.60**	-.20	.34	.57**	.51**	.42*	.43*	.12	.56**	.07	-.01	.25	.65**	-
MS	9.37	-2.65	-2.29	-4.54	-9.37	23.04	981.00	5.76	5.76	8.92	7.76	6.76	21.32	16.80
SD	2.42	2.86	2.93	3.83	6,33	9.39	336.01	.93	.93	1.96	1.74	.97	4.64	3.79

** p < .01. * p < .05

Effect approach analysis predictions from specific competencies on overall performance on WPs

We predicted that item characteristics relate to individual characteristics and that individuals with higher abilities in one domain were less affected by more complexity in that domain. To test this hypothesis, we have computed multiple step regressions of different individual capabilities for the response time on the consistency effect, on the nominalization effect, the carry effect, and the operation. Measures of individual capabilities were calculation skills, reading speed, reading comprehension (verbal as well as letters-span) and Corsi block backwards and forwards; these were included in the regression analyses as predictors.

For the **consistency effect** the regression model [R= .69, adjusted R^2= .22, F(1, 23) =1.87,p=.137] turned out not to be significant.

For the **nominalisation effect** the final regression model [R= .44, adjusted R^2= .16, F(2, 22) = 10.42, p=.001] included the following predictors: (i) reading comprehension, (ii) Corsi block forward test. The positive beta indicates that a relatively higher score in the reading comprehension (β=0.471) task was associated with a lower nominalisation effect, and the negative beta score of the Corsi block forwards (β=-0.421) indicated that a higher score was associated with higher nominalisation effect.

For the **operation effect** the final regression model [R= .57, adjusted R^2= .292, F(1, 23) =10.89,p=.003] included only the predictor, speed test subtraction. The positive beta (β=0.567) indicates that a relatively higher score in the subtraction speed calculation task was associated with a lower operation effect.

For the **carry effect** the final regression model [R= .51, adjusted R^2= .26, F(1, 23) =8.04,p=.009] included only the predictor, speed test subtraction. The positive beta (β=0.509) indicates that a relatively higher score in the subtraction speed calculation task was associated with a lower carry effect.

Discussion of the influence of individual capabilities

To examine whether characteristics of the individual also influence WP-solving we assessed reading comprehension, reading speed, mathematical abilities and working memory. The correlations in the Table 2 seem to hint that people with higher reading and mathematical abilities perform better when solving WPs: Numeracy and literacy skills had a significant negative correlation with reaction time and turned out to be significant predictors in the regression as well. However, surprisingly, individual working memory capability had no connection to reaction time in our study. This is unusual because many past studies (Adams & Hitch, 1997; Furst & Hitch, 2000; Passolunghi & Siegel, 2001; Swanson, 2004) have shown a strong influence of working memory on mathematical performance, and only a few studies did not find working memory to be a significant predictor (Fuchs et al., 2015). There are at least five reasons for the surprising finding that working memory did not predict WP performance. First, the method of presentation might also increase the working memory demand. In the present study, the WPs were present until the answer was submitted, and participants could reread the problems when desired. Second, it is possible that we have not seen an effect for working memory in adults because it may play a different role, e.g. the extra information should be less distracting than for children. Third, most studies have been conducted in children, while we tested adults. Children vary more widely in reading and other cognitive skills than adults (Kingsdorf, Krawec, & Gritter, 2016). Thus, the divergence in findings may be just a matter of working memory variance within the tested sample. Fourth, different assessments of working memory may lead to divergent results as different measures of working memory are often poorly correlated. Fifth, in the current study we have used one-step WPs, which might have led to lower cognitive load. According to Peng et al. (2016), multistep mathematical tasks that require the calculation and maintenance of intermediate values are hypothesized to draw more on working memory resources than mathematical tasks that consist of fewer steps. Thus, there are multiple potential reasons why individual working memory capability predicted performance in some studies, but not ours.

Influence on linguistic and arithmetic effects

The hypothesis that participants with better mathematical skills should be less affected by arithmetic complexity, and participants with better reading skills should be less affected by the linguistic complexities could be partially confirmed.

Arithmetic Effects. For both arithmetic effects, namely the operation effect (explained variance: 29%) and carry effect (explained variance: 26%) the final regression model included only the predictor, speed test subtraction. This suggested that a relatively higher score in the subtraction speed calculation task was associated with a lower effect size, i.e. more arithmetically more capable individuals had less difficulties with the more complex arithmetic conditions. Furthermore, the carry effect correlated positively with reading comprehension and reading speed. This is especially interesting because carry/borrow is an arithmetic factor unrelated to linguistics factors, yet we can still see correlations with reading comprehension and reading speed.

Linguistics Effects. In the case of the nominalization effect – linguistics factor unrelated to arithmetic – the final regression model (explained variance: 16%) included the predictor reading comprehension. Higher scores in reading comprehension lower the nominalisation effect, i.e., better readers had less problems with more complex grammatical structures. However, a high score in Corsi block forward was associated with a higher nominalisation effect. This result is surprising, but more working memory capacity can have disadvantages and lead individuals to employ complex strategies in problem-solving that are less optimal for a given task (DeCaro & Wieth, 2016). For the consistency effect, no individual skill turned out to be a significant predictor.

Limitations. Again, we wish to note that the N was rather small for our correlational and regression-based analysis. Some results, and especially some null results, may be due to the low power of this study. Therefore, we report these results only in this SOM, but not in the main paper. Although they generally correspond to the a priori hypothesis, they should be interpreted with care.

Summary

Individual reading and arithmetic abilities, but not working memory capacity, predicted overall performance. In effect-based analyses, in general, individual arithmetic capabilities predicted the arithmetic effects and individual linguistic and working memory capabilities predicted one, but not the other linguistic effect. Therefore, the data are consistent with the idea that arithmetically complex WPs are particularly difficult for students with poor arithmetic capabilities and linguistically complex WPs are particularly difficult for students with poor reading skills. With all due caution because of the low N, the data suggest that it could be promising to tailor the different linguistic and arithmetic complexities of WPs to the individual arithmetic and literacy capability of an individual

Study 3: Influence of Task Characteristics on Eye-Movement Patterns Related to Numerical and Textual Information in Arithmetic Word Problems

CONTRIBUTIONS OF CO-AUTHORS AND OTHER PERSONS

Authors

Gabriella Daroczy[1,3], Christina Artemenko[1,3], Detmar Meurers[2,3], Magdalena Wolska[3], & Hans-Christoph Nuerk[1,3]

[1]Department of Psychology, University of Tuebingen, Tuebingen, Germany

[2]Department of Linguistics, University of Tuebingen, Tuebingen, Germany

[3]LEAD Research Network, University of Tuebingen, Tuebingen, Germany

Author	**Author position**	**Scientific idea (%)**	**Data generation (%)**	**Analysis and Interpretation (%)**	**Paper writing (%)**
Daroczy, G.	1	70	100[2]	70	65
Meurers, W. D.	2	10	0	10	10
Artemenko, C.	3	0	0	0	15
Wolska, M.	4	5	0	5	0
Nuerk, H. C.	5	15	0	15	10

Proof-Reading

Zoë-Lauren Kirste

Student Research Assistants

Hanna Lönneker and Johanna Zeibig helped in data collection.

I/We certify that the above-stated is correct.

28.09.2020 Hans-Christoph Nuerk

Date, Signature of at least of one of the supervisors

[2] Responsible, students research assistants helped in data collection (50-60%)

ABSTRACT

Both linguistic and arithmetic task characteristics contribute to the difficulty of a word problems. However, the role of these characteristics and the exact cognitive processes underlying arithmetic word problems are often not clear, but they might be detectable by analysing eye-movement patterns. Not much is known about how eye-movements change under different linguistic and arithmetic task characteristics in regard to the whole problem as well as to specific parts of it (numerical and textual elements). This study examined the effects of linguistic and arithmetic task characteristics on the word problem-solving ability of children aged 10-13 years while their eye-movements were monitored. We manipulated the task characteristics independently of each other, using the mathematical factor operation (addition/subtraction) and the linguistic factors consistency (consistent/inconsistent) and nominalization (verbalized/nominalized). The results showed that eye-movements generally increase with increasing linguistic (i.e., nominalization) or arithmetic (i.e., operation) difficulty. However, specific parts of the text were differentially affected based on task characteristics: In general, increasing arithmetic difficulty shifts eye-movements towards numerical elements and increasing linguistic difficulty shifts eye-movements towards textual elements. However, the increase of difficulty in the arithmetic domain can also affect processing in the linguistic domain. For instance, as textual parts of the word problem were more frequently attended to when arithmetic difficulty increased but not vice versa. This indicates that in the process of word problem-solving, text comprehension and calculation are not sequential independent processes, but partially rely on the same processing components, such as working memory resources.

Keywords

eye-movement, arithmetic word problem, nominalization, lexical consistency, addition, subtraction

INTRODUCTION

Behavioural studies have shown that both linguistic and arithmetic task characteristics contribute to the difficulty of a WP (Barbu & Beal, 2010; Boonen et al., 2013; Daroczy et al., 2015; Davis-Dorsey et al., 1991; De Corte & Verschaffel, 1987; De Corte et al., 1985; De Corte et al., 1990; Lewis & Mayer, 1987; Orrantia & Múñez, 2013; Orrantia, Múñez, San Romualdo, & Verschaffel, 2015). However, relatively little is known about the exact relationship between linguistic and arithmetic task characteristics and the process of WP-solving. The key question is how people read (initial reading process), understand and create a mathematical model (i.e., integration and transition process which results in a mental model) and choose a calculation, which in the end they carry out (calculation process) for different types of WPs.

In research and modelling of WP-solving there is a debate about the existence or non-existence of common processing stages in the problem-solving processes: some suggest a (sequential) separation of the cognitive processes of reading and mathematical problem-solving (Fuchs et al., 2008; Fuchs, Gilbert, Fuchs, Seethaler, & Martin, 2018; Powell & Fuchs, 2014), while others argue for joint processing stages between the processes (Bergqvist & Österholm, 2010; Dehaene & Cohen, 1995). Whether the problem-solving processes are distinct or rely on joint processes can be reflected in the interaction of linguistic and arithmetic task characteristics (Daroczy et al., 2020).

In behavioural studies not all interaction effects between linguistic and arithmetic task characteristics can be detected because reaction times may be the same for different factors but do not reflect which parts of the text required greater attention, and thus play an important part in creating the mathematical mental model. Even if there is an interaction, we do not know whether it comes from arithmetic influences on linguistic factors and/or vice versa. On the other hand, with other methods, like eye tracking, we can get a deeper insight into the processes underlying linguistic and arithmetic difficulties and about the WP-solving process (Hegarty et al., 1992; Van der Schoot et al., 2009). For example, it is assumed that the elements in a WP that are looked at longer are central to the construction of the representation (i.e., mental model) and production of the solution (Van der Schoot et al., 2009). Monitoring eye-movements can provide this information. Therefore, the aim of the current study was to investigate the contribution of linguistic and arithmetic task characteristics using eye tracking in order to understand the direct influence of the task characteristics on the process of WP-solving as well as the possible common processing phases.

The direct influence of textual difficulty has already been shown to be visible in eye-movement data. Namely, increasingly difficult WPs elicit greater and more prolonged attention as well as re-readings of the overall problem (Hegarty et al., 1995; Strohmaier, Lehner, Beitlich, & Reiss, 2019). Additionally, linguistically complex WPs seem to increase the frequency and length of regressions to the number, relational term and variable names, as compared to linguistically less complex problems (Van der Schoot et al., 2009; Verschaffel et al., 1992). In this study, we expected to replicate these results.

Less is known about how task characteristics influence specific parts of a text. It has already been shown that highly capable students differ in looking at the numbers and keywords; and other part of the text compared to students who solve WPs with less ACC (De Corte et al., 1990; Hegarty et al., 1992; Van der Schoot et al., 2009; Verschaffel et al., 1992). Nevertheless, students who display less ACC were also able to apply successful strategies to solve WPs with low semantic difficulty and showed similar eye-movements for specific parts of the problems as compared with highly capable students (Van der Schoot et al., 2009). This might indicate that task characteristics elicit an attentional shift in distinct eye-movement patterns targeting specific parts of the WP in general; however, this has never been investigated for change problems, that is, problems that usually start with an initial quantity and require an action that causes either an increase or decrease in that quantity (Riley et al., 1983).

Lastly, this study also aimed to explore how the interaction of task characteristics manifests in eye-movement behaviour with the objective of investigating the separability of linguistic and arithmetic processes. Existing WP-solving models – the propositional theory (e.g., Kintsch & Greeno, 1985) and the cyclic model (e.g., Bergqvist & Österholm, 2010) – focus heavily on how the language (especially the first reading process) influences mathematical problem-solving. Yet, the models do not address if and how arithmetic processes affect the mental model. Eye-movements over numbers were not analysed separately from eye-movements over keywords in the majority of the previous studies, especially in those investigating the process of problem-solving and the creation of the mental model (Hegarty et al., 1992; Hegarty et al., 1995; Van der Schoot et al., 2009). In order to understand the relation of calculation and comprehension processes in WP-solving, it is especially important to separate areas of interest for numbers, keywords and other textual elements. Our hypothesis was that if linguistic and arithmetic processes are fully separable, increased linguistic complexity should lead to greater attention to the linguistic part of the problem only and increased arithmetic complexity should lead to greater attention to only the arithmetic part (i.e., the numbers) of the problem. On the other hand, if there is an interaction the linguistic/ arithmetic complexity should lead to different eye-movement patterns for both linguistic and arithmetic parts of the problem. The interaction could be uni- or bidirectional, for example, linguistic complexity could influence arithmetic processing or arithmetic complexity could influence linguistic processing.

In sum, very little is known about how task characteristics – i.e., different levels of linguistic and arithmetic difficulty – influence attention on WPs as indexed by eye-movement patterns as regards the whole problem or important parts of the text. This is very important because eye-movement analysis can detect both direct influences and interactions of the problem-solving processes. In the following paragraphs, we elaborate on the important linguistic and arithmetic task characteristics of WPs and lay out which ones we chose to examine in this eye tracking study with children.

ARITHMETIC TASK CHARACTERISTICS OF WORD PROBLEMS

Arithmetic factors influencing the difficulty of WPs are for instance, arithmetic operation, the types of numbers and calculation (De Corte, 1988), the number of calculations (Terao et al., 2004) and the magnitude of the numbers (Thevenot & Oakhill, 2005). Often, such arithmetic factors are neglected in eye tracking studies on WPs, but when they are systematically manipulated they usually influence performance. An arithmetic factor, which changes the underlying mathematical structure of a WP, is arithmetic operation. For instance, subtraction is considered to be more difficult than addition (Artemenko et al., 2018) and can be solved by different strategies (e.g., by indirect addition or direct subtraction Torbeyns et al., 2009). Operation is a relevant factor for investigating arithmetic processes and their relation to other problem-solving processes.

LINGUISTIC TASK CHARACTERISTICS OF WORD PROBLEMS

Linguistic factors, just like arithmetic factors, can influence the difficulty of a WP. Lexical consistency concerns specific keywords in the text (Hinsley et al., 1977). These keywords can be either consistent or inconsistent with the operation required for the correct solution. To address any incongruency between the text and the correct solution (i.e., for inconsistent problems), there is a need for different strategies than for consistent problems (Boonen et al., 2016). There is empirical evidence for the consistency effect (i.e., longer reaction times and higher reversal error rates and errors resulting from the choice of the opposite operation) for WPs containing inconsistent compared to consistent keywords (Hegarty et al., 1992; Hegarty et al., 1995; Pape, 2003; Van der Schoot et al., 2009; Verschaffel et al., 1992). One main goal of eye tracking research and WPs has been to locate the origin of the consistency effect (Hegarty et al., 1995; Verschaffel et al., 1992) and to postulate in which phase of the solution process the consistency effect occurs. It is hypothesised that the additional processing required for inconsistent problems, as compared to consistent problems, occurs not in the first reading period but during later phases of the solution process. Also, interactions between consistency and other factors play an important role. For instance, it has been shown that consistency interacts with operation, likely due to semantic complexity (Van der Schoot et al., 2009). In eye tracking studies involving WPs one way to increase the semantic complexity of (inconsistent) compare problems is to manipulate the relational term – i.e. markedness. Markedness refers to a central concept of linguistics in which one linguistic element is more distinctively identified (marked) than another (unmarked) element; however, the linguistic factor markedness cannot be independently manipulated from the factor operation (Van der Schoot et al., 2009). Because the factors markedness and operation are confounded, the independent analysis of eye-movement patterns for these factors as well as for their interaction is not possible. Therefore, in this study we needed a factor that increased semantic complexity independently of lexical consistency.

Even simple linguistic changes to the text can increase the difficulty of problem comprehension (Boonen et al., 2013; Van der Schoot et al., 2009). Contrary to markedness, some of the above-mentioned factors can be manipulated independently from operation and lexical consistency. Here, we use the factor nominalization, (a syntactic form that turns verbs into nouns, e.g., the nominalised form of the verb "to earn" is "the earning") which leads to longer reading times (Gibson & Warren, 2004). Nominalization has been shown to influence the reading difficulty of WPs (Prediger et al., 2015) but does not affect the underlying arithmetic structure of the problem. As mentioned before nominalization, contrary to markedness, serves to increase semantic difficulty. Since nominalization is related to reading difficulty it therefore can be associated more with the initial reading phase.

There is agreement in the literature that both linguistic and arithmetic attributes contribute to WP difficulty (for a review see Daroczy et al., 2015) but the studies provide contradictory evidence concerning the relation between linguistic and arithmetic factors. While the interaction between linguistic and arithmetic factors has been shown to manifest in the solution process (De Corte et al., 1990; Hegarty et al., 1992; Van der Schoot et al., 2009; Verschaffel et al., 1992), it has not yet been shown in eye-movement behaviour.

EYE-MOVEMENT IN WORD PROBLEM RESEARCH

A general assumption in eye-movement research and mathematics is that more numerous and difficult stimuli lead to longer fixations (Knoblich, Ohlsson, & Raney, 2001). Fixation time reflects the sum of the gaze time devoted to the areas of interest. This has also been observed in the case of WPs: students spent more fixation time on the whole problem for complex problems than for simple problems (De Corte et al., 1990). Another important measure for the duration of eye-movements, which has been applied in the domain of WPs, is regression time, where people look back at previously fixated areas (Frisson & Pickering, 1999; Rayner, 1998; Van der Schoot et al., 2009). Refixations and regressions are believed to often result from comprehension issues because additional processing for more difficult problems is assumed to occur during phases after the first reading phase (De Corte et al., 1990; Hegarty et al., 1992; Kintsch & Greeno, 1985; Verschaffel et al., 1992), and is associated with putting more effort into building a problem model (Strohmaier et al., 2019). Additionally, in the study of Strohmaier et al. (2019) global measures, for example mean fixation time, while considering the problem as one text, were associated both with task difficulty and abilities of the reader. Therefore, we would expect that more difficult task characteristics elicit longer fixation times on the whole problem and a greater frequency of regressions to previously seen areas of the problem.

Much less is known about how linguistic and arithmetic task characteristics generally guide attention to various parts of the text. The assumptions made by studies on reading also hold for studies on arithmetic WP-solving: the elements in a WP that are fixated upon longer are assumed to be more deeply processed and, consequently, central to the construction of the representation and production of the solution (Van der Schoot et al., 2009). For example, linguistically more complex WPs are hypothesised to induce longer and more numerous regressions to the numbers, relational terms and variable names than linguistically less complex problems (Van der Schoot et al., 2009; Verschaffel et al., 1992). Additionally, Moeller, Klein, et al. (2011b) found that carry problems were associated with a longer total reading time and a greater number of regressions than non-carry problems. Therefore, in this current study we expected increased linguistic difficulty to shift eye-movements mostly towards the (difficult) textual elements, and increased arithmetic difficulty to shift movements mostly towards the (difficult) numerical elements, as these parts require more attention. This hypothesis has also been proposed to explain the results of previous behavioural studies (e.g., Daroczy et al., 2015, for a review), but whether or not it is true (as indexed by eye-movements) has never been tested before.

If the cognitive processes of reading and calculation are fully separable for WPs, increased linguistic difficulty should lead to increased attention towards the textual part of the problem only and arithmetic processes on the numerical part of the problem only. Nevertheless, if linguistic and arithmetic difficulty interact or rely on a common processing phase, they should lead to a different eye-movement pattern also in those areas which correspond to the other part of the problem associated with the manipulation (i.e., linguistic elements for greater arithmetic difficulty and numerical elements for greater linguistic difficulty).

Linguistic factors do not only influence linguistic difficulty, but also interact with arithmetic factors. One reason why linguistic difficulty should affect numerical areas is that the linguistic attributes of WPs can influence the difficulty of the calculation. For instance, Bassok et al. (1998) and Boonen et al. (2016) have shown that functionally connected objects (e.g., tulips–vases) are easier to add up in WPs than categorically related objects (e.g., tulips–daisies). This suggests that depending on problem formulation the calculation process might be different. Such an interaction should be especially pronounced in the case of lexical consistency (Verschaffel et al., 1992). If we see increased fixation time on numbers for inconsistent problems, it is evidence that in some cases complex linguistic factors make the calculation more difficult.

Arithmetic factors not only influence mathematical difficulty, but also interact with linguistic factors. As concepts of addition and subtraction (e.g., giving, taking) are evoked in everyday language (Terezinha N Carraher et al., 1988), they can be easily connected to linguistic features in WPs (Lave, 1992). Calculation failures are mostly hypothesised to result from the improper construction of a mental model for the WP, this would indicate that operation is involved in the building of a mental model, which, furthermore, is associated with greater attention and re-readings of the textual elements of the WP (Hegarty et al., 1992). This indicates that not only the numbers, but also the textual elements of a WP could be affected by a change in the eye-movement pattern when the arithmetic difficulty increases.

Underlying Cognitive Processes. Investigating eye-movement behaviour can shed light on underlying cognitive processing (De Corte et al., 1990; Hegarty et al., 1992; Hegarty et al., 1995; Verschaffel et al., 1992). Many studies rely on the model of Kintsch and Greeno (1985) and assume that WP-solving requires four distinct phases (Mayer, 1984): an initial reading phase (translation of the text), integration (mental representation and the construction of the problem model), planning (generating a solution plan) and solution execution (calculation). If participants construct a model, they frequently focus on specific parts of the text while reading through the problem (De Corte et al., 1990; Hegarty et al., 1992; Van der Schoot et al., 2009; Verschaffel et al., 1992). The proposed models for WP-solving suggest differences in terms of the relationship between problem-solving phases. Some suggest a separation of the cognitive processes of reading and calculation (Fuchs et al., 2008; Fuchs et al., 2018; Powell & Fuchs, 2014), while others argue for an interaction (Bergqvist & Österholm, 2010; Dehaene & Cohen, 1995). According to the propositional theory model, the textual difficulty (here: nominalization) affects only the initial text comprehension processing and, also, arithmetic processing does not interfere with processing phases other than the calculation phase. However, other types of models contradict these assumptions and suggest common or interactive processing for linguistic and numerical information: For instance, Bergqvist and Österholm (2010) suggest that text formulation affects not only the first reading but also other phases in the solution process and that problem comprehension and computational processes interact. This means that the model assumes that we cannot separate the reading and solving of arithmetic tasks. The first reading creates a version of the mental representation. The authors also found that carrying out a calculation adds to the mental representation, while re-reading actually changes the existing mental representation.

The propositional, sequential and cyclic models suggest different eye-movement patterns, which we investigated in the present study. First, if the propositional theory was true, we expected a clear difference between the measure es regression time and fixation time for WPs, because fixation time also involves the initial reading phase, and regression time is associated with the phases after the initial reading period. This would mean that nominalization, which may be associated with the initial reading phase, should increase the total fixation time but not the regression time on the text or the numbers because it would only increase the reading difficulty. On the other hand, consistency, which is associated with the mental model building phase (Verschaffel et al., 1992), should affect the regression times. This is supported by findings that lexical consistency elicited more regressions on the other parts of the text than the keyword of the texts (Hegarty et al., 1992) According to the propositional model, only the factor operation should affect the numbers because it is involved in the calculation process. According to models which argue for the non-sequentially of problem-solving, nominalization should also affect the numerical areas. This suggests that the initial reading process influences another phase of problem-solving and is part of the mental representation phase. Similarly, if operation influences eye-movement behaviour over the text, text comprehension would also be confounded with the calculation process during WP-solving when the difficulty of the operation increases. In this way, eye-movements enable a deeper understanding of the (possibly interactive) influence of task characteristics on the process of WP-solving.

THE PRESENT STUDY

In this study, we aimed to investigate how linguistic and arithmetic difficulty affect the processing of the whole problem as well as specific parts of (i.e., numerical and textual elements). In this within-subject design we included one arithmetic factor and two linguistic factors orthogonally. The arithmetic factor was operation (addition vs. subtraction) and the linguistic factors were lexical consistency (consistency vs. inconsistency) and nominalization (verbalization vs. nominalization).

The first hypothesis was that longer fixation and regression times and lower ACC could be expected for the whole problem for generally complex items (subtraction, inconsistency, nominalization). This hypothesis is a conceptual replication of previous eye-movement studies using WPs and serves as a validity test. The second more specific hypothesis not previously tested as such was that with the increasing difficulty of arithmetic information eye-movements are shifted towards numerical elements and that with the increasing difficulty of linguistic information eye-movements are shifted towards textual elements. Moreover, we expected an interaction of linguistic and arithmetic factors on the total fixation time as well as on regression times on specific areas of areas of interests. Namely, for linguistically complex WPs (lexical inconsistency, nominalization), we expected longer fixations not only on the textual elements but also on the numerical elements of the WP. For arithmetically complex WPs we expected greater fixation time not only on the numerical elements but also the textual elements.

METHODS

PARTICIPANTS

In total, 42 children between 10 and 13 years old were tested. All children were native German speakers with normal or corrected-to-normal (only with soft contact lenses) vision and had no history of neurological or psychological disorders. The study was approved by the local ethics committee of the Medical Faculty of the University of Tuebingen and all procedures were in accordance with the latest version of the Declaration of Helsinki.

Out of the 42 children, 33 children (19 boys, 14 girls; age: $M = 11.48$, $SD = 0.66$ years) were included in the analysis. Participants were excluded because of technical problems ($n = 4$), incomplete data ($n = 2$), or an overall error rate above 40% ($n = 3$).

APPARATUS

We used a remote version of the SR Research EyeLink 1000 eye tracker (from SR Research Ltd, Ottawa, Ontario) that does not require a forehead/chin rest and uses a sampling rate of 500 Hz. The system tracks eye-movements by measuring changes in pupil position in a video image. The best eye to record was automatically selected during a nine-point calibration procedure. Calibration is the process used to set the eye-movement software to accurately track eye-movements. Calibration was accepted when the worst error in gaze position was smaller than 1.5° and the average error was smaller than 1.0°. Participants in the experiment were seated such that the distance between their eyes and the monitor (51 cm diagonally with a resolution of 1025 x 768), on which the WPs were displayed, was approximately 70 cm. The height of the chair was adjusted based on the stature of each child so that the line of sight was standardized across all participants.

STIMULI AND DESIGN

The study had a 2 x 2 x 2 within-subject design with the factors operation (addition/subtraction), consistency (consistent/inconsistent) and nominalization (verbalized/nominalized; see Table 2). Forty-eight simple one-step arithmetic WPs in German were designed for the study (see Supplementary Material 1), each consisting of 4 sentences. The first sentence was to be used as a reference point (e.g., *"A market woman is selling apples at the market."*). The second and the third sentences contained the two numbers necessary for the mathematical operation (e.g., *"The market woman arrived with 42 apples. She sold 17 apples.").* The fourth and final sentence contained the question (e.g., *"How many apples did she have at the end?*").

The problems were constructed in such a way that manipulation of one factor did not affect the other factors (i.e., the factors were manipulated orthogonally). All WPs belonged to the type *Change* according to the categorization of Riley et al. (1983). We also ensured that each task could be clearly solved with one calculation, that the text was not ambiguous and that there was only one correct solution. Finally, all sentences were constructed using active verb forms because the active voice helps people to better understand WPs (Abedi et al., 2005; Shaftel et al., 2006). Unnecessary information was also avoided because the presence of extraneous information reduces student solution ACC (Muth, 1992).

All problems included two two-digit numbers in Arabic notation as operands constituting addition or subtraction as the arithmetic operation. Overall problem size and parity were matched between subtraction and addition. Pure decades (e.g., 50), tie numbers (e.g., 66), unit ties (e.g., 23 - 13), decade ties (e.g., 41 + 43), and mirror numbers (e.g., 24 - 42) were excluded to prevent any automatic mental retrieval. We kept carry operations constant (i.e., all addition tasks needed carrying and all subtraction tasks needed borrowing), because we wanted not only text understanding, but also arithmetic computations to be demanding (see: Artemenko et al., 2018; Moeller, Klein, et al., 2011b).

Lexical consistency was manipulated by systematically varying both the keyword and the required arithmetic operation: The second or third sentence contained the keyword, which evoked an operation (e.g., "sell"/"selling" for subtraction, "buy"/"buying" for addition). The WP, however, was constructed such that both operations, addition and subtraction, were equally likely for each keyword type, resulting in trials that were 50% lexically consistent and 50%

inconsistent. The linguistic factor nominalization was manipulated independently of consistency and operation: The type of nominalization in the text was formulated so that the words lost their verbal characteristics and behaved like real nouns (Hamm & van Lambalgen, 2002). All nominalized forms originated from the same lexical root as the corresponding verbs, illustrated by the following example: Nominalized form: *"He* ***earns*** *15 euros."* Verbalized form: *"He is happy about* ***the earning*** *of 15 euros."*

PROCEDURE

The experiment took place in a laboratory of the University of Tuebingen. All participants and caregivers were first informed about the nature and structure of the experiment. After signing the informed consent, the participants filled out questionnaires for neuropsychological measurements: arithmetic ability, intelligence and reading comprehension. Arithmetic ability was assessed by two speed calculation tests on addition and subtraction (Huber et al., 2013) that looked at the number of correctly solved items in 90 s for two-digit addition and two-digit subtraction. Intelligence was assessed by the subtest Matrices from the HAWIK (Petermann & Petermann, 2008). Reading ability was assessed by the LGVT test (Schneider et al., 2007) with the sub-scales reading speed (i.e., the number of read characters) and reading comprehension (i.e., correctly solved reading comprehension items). Verbal working memory was assessed with letter span forward and letter span backward (Individual Characteristics and Eye-Movement Patterns: see Supplementary Material 4).

For the experiment, subjects were instructed that only mental calculations were allowed and that no other calculation support (such as finger movements) should be used. To ensure that all participants started to encode the addition problem at the first summand, a fixation point to the left of this summand was used (x/y coordinates: 112/384 pixel). Vocal response was used to detect the reaction time. The Voice Key (Creative EMU 0202, USB audio interface, Scotts Valley, USA) was used as a trigger, consisting of the interface and a headset with a microphone. Voice Key continuously monitors the voice volume and triggers a recording as soon as a user-defined limit is exceeded. Past research has shown, however, that there may be a conditional bias in recording louder sounds using Voice Key so, for instance, phonemes in the beginning might be registered with a delay (Kessler et al., 2002; Tyler, Tyler, & Burnham, 2005). For the Voice Key calibration, 32 two-digit numbers were each presented centrally on the screen in two blocks, and participants had to loudly read the numbers. Following Voice Key calibration, the main experiment started. In the eye-movement experiment, there was a fixation point at the beginning of each trial.

The stimuli were presented on a 19-inch monitor in white on a black background (Times New Roman font, font size 24, bold, all four sentences at an equal distance from one another). To reduce potential measurement inaccuracy or at least to keep it constant, all participants were asked to say loudly and clearly "is" before pronouncing a number (e.g., "ist 45") in both the Voice Key calibration and the main experiment. As soon as the participants responded the trial disappeared, and the screen turned black for 360 ms. There was a time limit of 90 s for response. Participant responses were recorded by the experimenter. The 48 trials were divided into two blocks with 24 trials each with a break in between. The trials were presented separately in a pseudo-randomized order, so that in each block all 8 conditions were included, the conditions were mixed, and a single condition was presented no more than three times in a row.

DATA ANALYSIS

Behavioural Data Analysis. Response time (RT) was defined as the time between the start of the WP presentation on the screen and the trigger of the participant's response sent by the Voice Key. For RT and accuracy (ACC) analyses, some responses were excluded due to errors in the Voice Key: (1) Voice Key was not triggered, so that the solution had to be repeated, (2) the subject corrected the answer after the first response, (3) other sounds triggered the Voice Key (e.g., throat clearing). For RT analysis, incorrectly solved trials were excluded. Additionally, all RTs shorter than 3,000 ms and longer than 30,000 ms were excluded. Furthermore, a trimming procedure eliminated all RTs falling below or above three SDs of an individual participant's mean. For statistical analysis of RT and ACC in WPs, 2 x 2 x 2 ANOVAs were conducted with the factors: operation (addition/subtraction), consistency (consistent/inconsistent) and nominalization (verbalized/nominalized).

Eye-Movement Analysis. Following the method of Hegarty et al. (1995) and Van der Schoot et al. (2009), regression times and fixation time were calculated. Fixation time was defined as the total fixation time on an area. The regression times were defined as the total fixation time on an area minus the first pass time (Frisson & Pickering, 1999; Rayner, 1998). This includes all regressions after the first pass fixation period. First pass duration on the target areas of interest was defined as the sum of the duration of all fixations on the areas of interest from the first fixation on the areas of interest until the first time that the reader left the areas of interest. There are also other eye-movement measures such as fixation and regression count, but since, for example, the pattern of the regression count corresponds to the pattern of regression time in WP research (Van der Schoot et al., 2009), we focused our analyses only on fixation time and regression time.

The eye-movement measures fixation time and regression time were calculated for the following areas in each WP (see Figure 1[3]): the total text (I0), the numbers (I1), the total text minus the numbers (I2) and the keyword (I3).

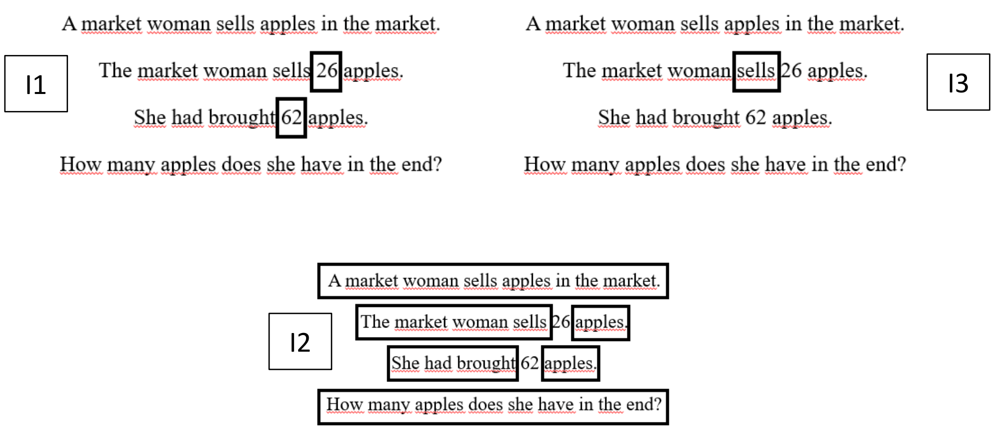

Figure 3 - 1 Combined Areas of Interests (CAOIs) for Numbers (I1), Text without Numbers (I2) and Keyword (I3).

In order to handle these discontinuous areas, we calculated combined areas of interests (CAOIs). The basis of the calculations were fixations on individual words in the areas I0-I3. In the technical sense each word has its own area of interest (AOI). The first pass time for the larger CAOIs are the sum of the first pass time of the word AOIs belonging to the CAOI. The regression time for the larger CAOIs are the sum of the total fixation time of the word AOIs minus the sum of the first pass time of the word AOIs belonging to the CAOI.

Additionally, we have chosen absolute measures (mean fixation time, mean regression time) instead of relative measures (fixation time per number of words, percentage of regression time) for two reasons. The first reason is that we aligned ourselves with studies, for instance, Van der Schoot et al. (2009), that used absolute measures. The second reason is that our stimuli were comparable in length, word count and frequency, contrary, for instance, to studies like that of Strohmaier et al. (2019), where the stimuli differed more and relative measures were necessary. Table 1 provides the average word count, character length, syllable count, character count per word and syllable count per word for each condition as well as frequency measures.

[3] Throughout the paper, for presentation purposes we show English word-by-word glosses of the German example stimuli that were used. The original German stimuli can be found in the Supplementary Material 1.

Frequencies were calculated using the COW German Corpus (Schäfer, 2015; Schäfer & Bildhauer, 2012).

Table 3 - 1 *Linguistics Measures across Conditions.*

	Word Count		Character Length		Syllable Count		Character Count / Word		f_logpermil_10		band	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Addition												
Consistent Form												
Nominalised Form	30.17	4.07	144.67	27.04	47.33	9.73	4.64	0.35	121.30	10.81	7.13	0.42
Verbalised Form	24.67	2.73	117.33	18.64	38.33	7.28	4.58	0.34	115.71	10.17	7.13	0.53
Inconsistent Form												
Nominalised Form	29.00	4.20	144.33	29.19	46.50	10.13	4.84	0.39	120.42	9.55	7.37	0.27
Verbalised Form	25.50	3.33	126.50	25.26	40.17	9.33	4.77	0.39	114.11	8.43	7.43	0.50
Subtraction												
Consistent Form												
Nominalised Form	29.50	2.35	147.50	15.85	45.67	5.35	4.92	0.35	120.62	7.98	7.42	0.22
Verbalised Form	26.00	1.67	128.17	15.04	41.00	4.94	4.76	0.34	117.58	6.63	7.33	0.52
Inconsistent Form												
Nominalised Form	30.17	2.64	144.00	16.04	45.83	5.91	4.77	0.38	122.18	11.87	7.31	0.59
Verbalised Form	26.50	2.59	130.67	21.41	43.00	8.07	4.76	0.51	115.83	11.10	7.29	0.54

Contrary to some eye-movement studies (Hegarty et al., 1992; Hegarty et al., 1995; Van der Schoot et al., 2009; Verschaffel et al., 1992) that chose to analyse only the words referring to one of the quantities in the problem, we decided to include all of the words of the WP in the analysis. Because, according to Hegarty et al. (1992) students using successful solution strategies pay attention to all words in the problem and not only to the single words which refer to one of the quantities.

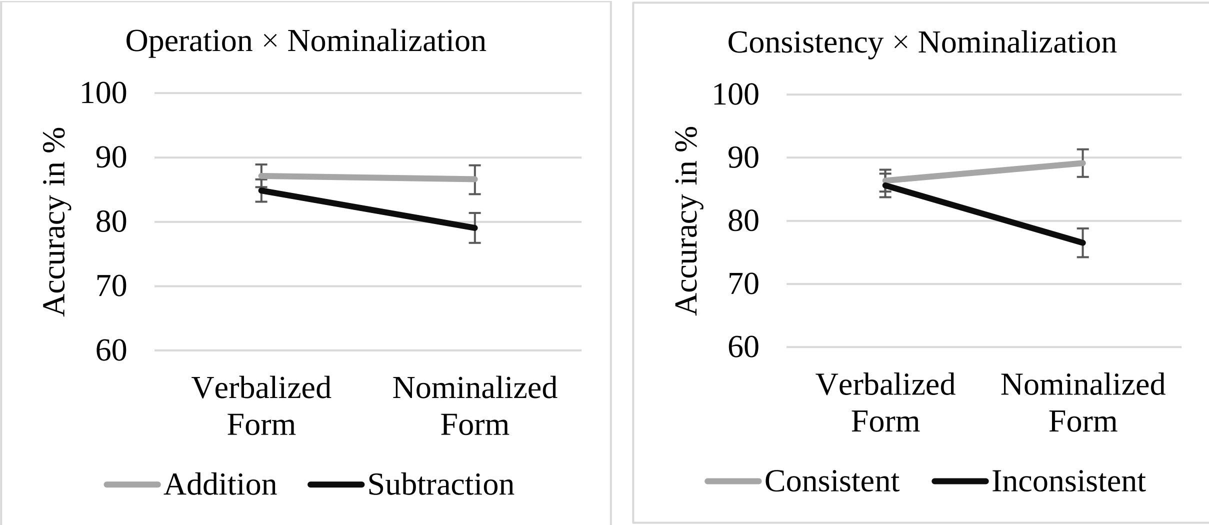

Figure 3 - 2 Two-Way Interactions in Accuracy: Operation × Nominalization, Consistency × Nominalization.

For statistical analysis of the areas of interests, 2 x 2 x 2 ANOVAs were conducted with the factors operation (addition/subtraction), consistency (consistent/inconsistent) and nominalization (verbalized/nominalized) separately for the following CAOIs: whole problem (I0), numbers (I1), whole problem without numbers (I2) and keyword (I3). To understand the consistency effect, we also examined the keyword separately from the numbers, because according to Experiment 3, Verschaffel et al. (1992), there was a significant effect of language consistency only for the second sentence which included the relational statement. These results indicate that the consistency effect manifests in the sentence where the keyword is included.

Finally, as numbers and keywords were considered areas of interest in most previous studies (e.g., Hegarty et al., 1992), an additional analysis of the areas of interest including keyword and numbers is included in the Supplementary Material 2. Furthermore, because the areas of interest including keyword and numbers and the areas of interest including the whole problem without keyword and numbers are associated with specific problem-solving strategies, the analyses of these areas can be found in the Supplementary Material 3.

RESULTS

BEHAVIOURAL MEASURES

Table 2 shows the statistical data for RT and ACC.

Table 3 - 2 *Descriptive Data for Behavioral Measures.*

	Verbalized Form								Nominalized Form							
	Consistent Form				Inconsistent Form				Consistent Form				Inconsistent Form			
	Add		Sub		Add		Sub		Add		Sub		Add		Sub	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD

Reaction Time (s)	21.00	1.55	24.65	1.54	21.75	1.55	25.89	1.77	23.76	1.61	26.56	1.98	22.54	1.32	27.03	1.91
Accuracy (Percentage)	88.38	1.98	84.34	2.61	85.86	2.91	85.35	2.69	91.41	1.94	86.87	2.69	81.82	2.75	71.21	3.50

Table 3 summarises the results from the ANOVAs of RT and ACC.

Table 3 - 3 *Analysis of Variance for Behavioural Data.*

	Reaction Time			Accuracy		
	F	*p*	η_p^2	*F*	*p*	η_p^2
Nominalization	9.43	.004	.23	2.94	.096	.08
Consistency	0.43	.519	.01	11.17	.002	.26
Operation	21.85	<.001	.41	4.87	.035	.13
Nominalization x Consistency	1.95	.172	.06	12.90	.001	.29
Nominalization x Operation	0.68	.796	.00	4.47	.042	.12
Consistency x Operation	1.21	.279	.04	0.10	.755	.00
Nominalization x Consistency x Operation	0.47	.496	.01	3.57	.068	.10

REACTION RIME

For RT (*M* = 23.93, *SD* = 8.43), there was a significant main effect of operation, indicating that subtraction problems took longer to solve (26.03 s) than addition problems (22.26 s). There was also a significant main effect of nominalization, indicating that WPs with the nominalized form took longer to solve (24.97 s) than problems with the verbalized form (23.32 s). All other effects and interactions were not significant.

ACCURACY

For ACC (*M* = 84.41, *SD* = 7.00), there were significant main effects of operation and consistency. Moreover, there was a significant interaction of *consistency and nominalization*, indicating a consistency effect (reduced ACC for inconsistent problems (76.5%) compared to consistent problems (89.1%) only in nominalized sentences, $t(32) = 4.57$, $p < .001$, but not in verbalized sentences; cf. Figure 2). Additionally, there was a significant interaction of *nominalization and operation*, indicating an operation effect (reduced ACC for subtraction (79.04%) compared to addition (86.61%) only in sentences with the nominalized form, $t(32) = 2.83$, $p = .008$, but not in sentences with the verbalised form; cf. Figure 2). All other effects and interactions were not significant.

Summary of Results. For RT, effects for operation and nominalization were found. For ACC, effects of consistency and operation were found only for sentences with the nominalized form (i.e., semantically more complex sentences) but not for easier sentences with the verbalised form.

EYE-MOVEMENT MEASURES

Table 4 shows the statistical data for fixation time and regression time.

Table 3 - 4 *Descriptive Data for Eye-Movement Measures (Means and SD in s).*

	Verbalized Form								Nominalized Form							
	Consistent Form				Inconsistent Form				Consistent Form				Inconsistent Form			
	Addition		Subtraction		Addition		Subtraction		Addition		Subtraction		Addition		Subtraction	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Fixation Time																
Whole Problem (I0)	17.89	7.86	20.99	7.67	18.86	8.56	21.94	8.71	19.99	7.59	22.3	9.65	18.95	6.05	22.85	9.03
Numbers (I1)	6.54	4.21	8.1	3.56	6.5	3.08	8.33	4.17	6.39	3.74	7.9	3.68	5.5	2.39	8.38	4.57
Text without Numbers (I2)	11.35	4.27	12.89	5.08	12.36	6.40	13.61	6.35	13.6	5.44	14.40	7.43	13.45	4.69	14.47	6.21
Keyword (I3)	0.87	0.44	0.97	0.48	1.01	0.93	0.67	0.37	0.60	0.23	0.80	0.32	0.89	0.35	0.86	0.36
Regression Time																
Whole Problem (I0)	1.24	7.05	14.97	7.14	13.15	7.62	15.91	7.99	13.25	6.44	15.73	8.84	12.33	5.10	16.45	8.26
Numbers (I1)	5.79	4.05	7.27	3.54	5.67	2.97	7.48	4.06	5.57	3.62	7.01	3.56	4.71	2.31	7.60	4.53
Text without Numbers (I2)	6.62	3.55	7.70	4.62	7.48	5.59	8.43	5.60	7.68	4.33	8.72	6.64	7.63	3.82	8.85	5.38
Keyword (I3)	0.52	0.42	0.59	0.44	0.63	0.89	0.40	0.35	0.33	0.18	0.46	0.25	0.52	0.29	0.55	0.32

Table 5 shows the ANOVAs about fixation time and regression time on the defined areas of interests.

Table 3 - 5 *Analysis of Variance for Eye-Movement Measures.*

	Whole Problem (I0)			Numbers (I1)			Text without Numbers (I2)			Keyword (I3)		
	F	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Fixation Time												
Nominalization	5.23	.029	.14	1.76	.194	.05	13.33	.001	.29	1.66	.207	.05
Consistency	0.73	.400	.02	0.06	.811	.00	1.36	.253	.04	1.57	.219	.05

Operation	23.13	< .001	.42	20.32	< .001	.39	11.06	.002	.26	0.12	.729	.00
Nominalization x Consistency	1.64	.209	.05	0.64	.430	.02	1.05	.313	.03	9.27	.005	.22
Nominalization x Operation	0.00	.988	.00	1.3	.264	.04	0.59	.447	.02	5.27	.028	.14
Consistency x Operation	0.92	.346	.03	2.28	.141	.07	0.00	.960	.00	13.96	.001	.30
Nominalization x Consistency x Operation	0.97	.332	.03	1.44	.239	.04	0.18	.675	.01	1.25	.272	.04
Regression Time												
Nominalization	0.51	.481	.02	1.78	.192	.05	3.51	.070	.10	0.99	.327	.03
Consistency	0.87	.357	.03	0.04	.835	.00	1.71	.990	.05	2.3	.139	.07
Operation	21.54	< .001	.40	18.94	< .001	.37	10.28	.003	.24	0.00	.956	.00
Nominalization x Consistency	1.16	.290	.03	0.22	.639	.01	0.82	.373	.03	4.64	.039	.13
Nominalization x Operation	0.60	.444	.02	1.43	.241	.04	0.04	.843	.01	2.76	.107	.08
Consistency x Operation	1.30	.263	.04	2.63	.114	.08	0.00	.967	.00	6.21	.018	.16
Nominalization x Consistency x Operation	0.88	.356	.03	1.58	.218	.05	0.08	.784	.02	1.32	.259	.04

General combined area of interest. The whole problem (I0). In the case of *total fixation time* (M = 20.42, SD = 7.38), there were significant main effects of operation and nominalization. These indicate that subtraction problems took longer to solve (21.99 s) than addition problems (18.96 s), and problems with the nominalized form (20.92 s) took longer to solve than problems with the verbalized form (19.92 s). In the case of *total regression time* (M = 14.24, SD = 6.51), there was a significant main effect of operation, indicating that subtraction problems (6.68 s) elicited more regression time on the total text than addition problems (5.50 s). Thus, the general eye-movement data is consistent with earlier studies.

Specific combined area of interest: The two numbers (I1). For *fixation time looking at the numbers*, there was a significant main effect of operation, indicating that for subtraction participants had a longer fixation time (8.18 s) than for addition (6.33 s). For *regression time*, there was a significant main effect of operation. This indicates that for subtraction participants displayed more regression time (7.34 s) than for addition (5.43 s). No linguistic manipulation (consistency, nominalization) led to any effects regarding the fixation patterns on numbers.

Specific combined area of interest: Whole problem without numbers (I2). For *fixation time*, there were significant main effects of operation, indicating that subtraction leads to a longer fixation time (13.84 s) than addition (12.69 s), and of nominalization, indicating that nominalization (13.98 s) leads to longer fixations than verbalization (12.55 s).

For *regression time*, there was a significant main effect of operation, indicating that subtraction elicited a longer regression time (8.43 s) than addition (7.35 s). Therefore, not only linguistic, but also arithmetic factors influenced eye-movement effects in the text without numbers.

Specific combined area of interest: Keyword (I3). For *fixation time*, there was a significant interaction of nominalization and operation, indicating that there was a significant difference between the addition (0.73 s) and subtraction (0.82 s) sentences in the case of nominalization, $t(32) = -2.13$, $p = .004$, but not verbalization. Additionally, there was a significant interaction of nominalization and consistency, indicating a significant difference between inconsistent (0.87 s) and consistent sentences (0.69 s) in the case of nominalization, $t(32) = -5.44$, $p < .001$, but not verbalization. Finally, there was a significant interaction of consistency and operation, indicating longer fixation times for consistent sentences (0.89 s) than inconsistent sentences (0.75 s) for subtraction, $t(32) = 2.56$, $p = .015$, but longer fixation times for inconsistent sentences (0.95 s) than consistent sentences (0.73 s) in the case of addition, $t(32) = -3.73$, $p = .001$ (cf. Figure 3).

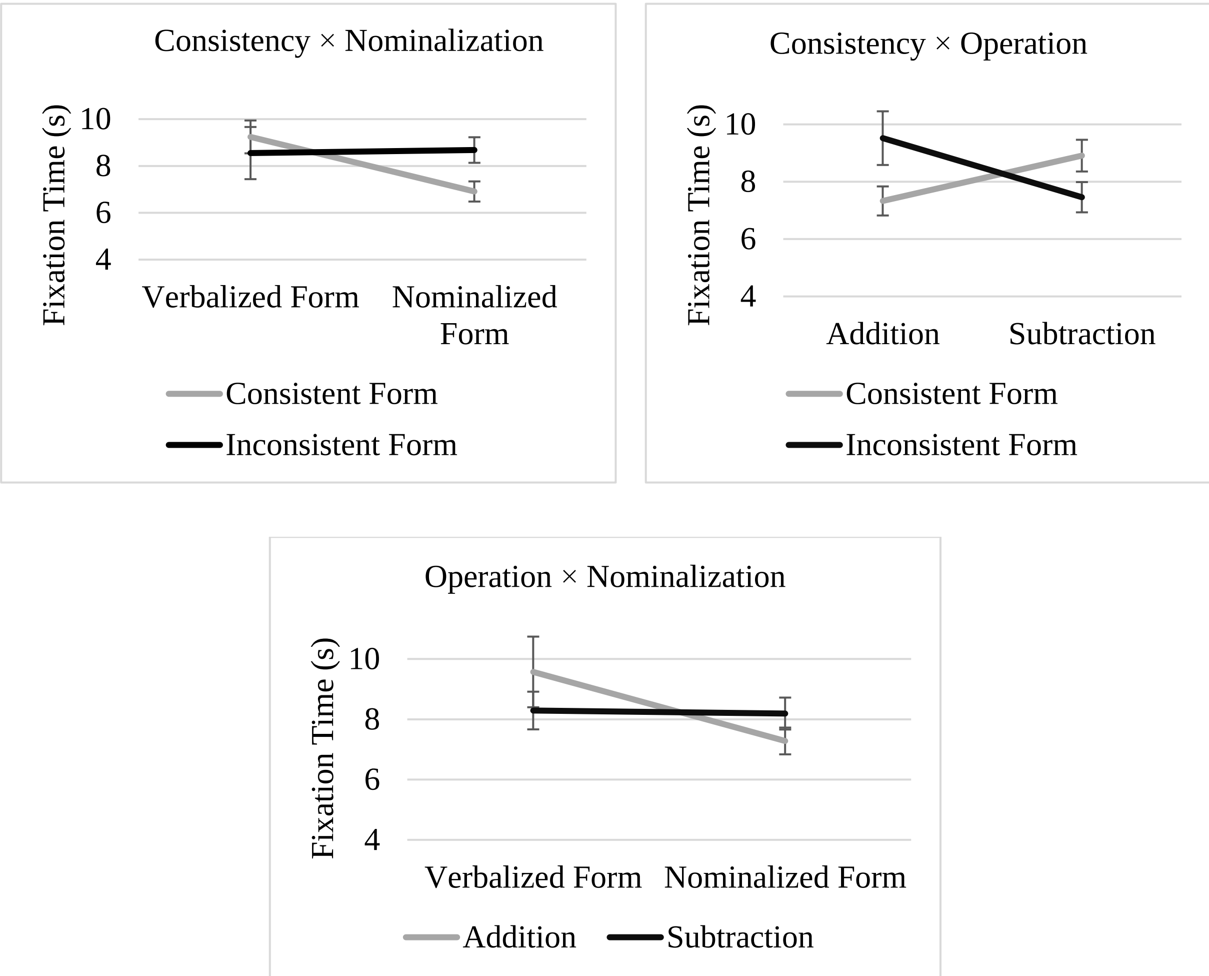

Figure 3 - 3 Two-way Interactions in Fixation Time on Keyword (I3): Consistency × Nominalization, Consistency × Operation, Operation × Nominalization.

For *regression time*, there was a significant interaction of nominalization and consistency, indicating a significant difference between inconsistent (0.53 s) and consistent sentences (0.39 s) in the case of nominalization, $t(32) = -4.45$, $p < .001$, but not verbalisation (cf. Figure 4). Furthermore, there was a significant interaction between consistency and operation, indicating a significant difference between consistent (0.42 s) and inconsistent (0.58 s) sentences in the case of addition, $t(32) = -2.97$, $p = .006$, but not in the case of subtraction (cf. Figure 4).

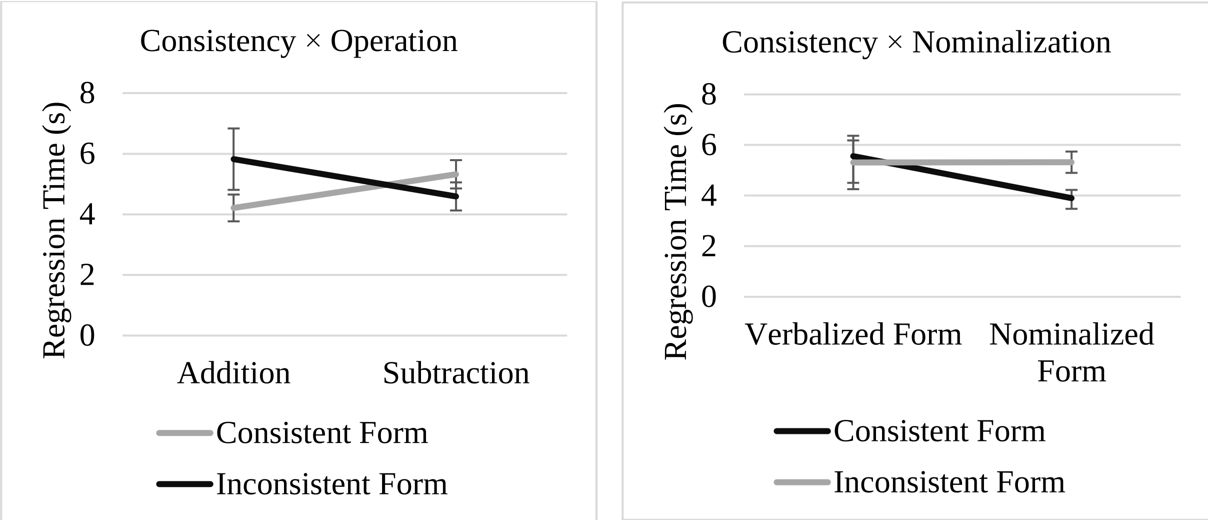

Figure 3 - 4 Interactions in Regression Time on Keyword (I3): Consistency × Operation, Consistency × Nominalization.

In sum, both linguistic and arithmetic factors as well as their interaction influenced eye-movement patterns over keywords.

Summary of Results. For the whole problem including numbers (I0), effects for operation and nominalization were found. Fixation and regression time on numbers (I1) provided effects for operation, i.e. only for numerical manipulations. In contrast, fixation and regression time on text without numbers (I2) provided an operation effect as well as a nominalization effect for fixation time only. Thus, eye-movement behaviour over the text *without numbers* is not only influenced by linguistic complexity, but also by numerical complexity. This holds even for eye-movement data for one single word, the so-called keyword. Fixation and regression time on the keyword (I3) provided an interaction of the linguistic factors, namely, a consistency effect only in case of nominalization. However, it also provided evidence for interaction of linguistic and arithmetic factors, namely between nominalization and operation as well as consistency and operation. The interaction on the keyword (I3) between consistency and operation was disordinal.

DISCUSSION

In this study we systematically investigated the relationship between linguistic and arithmetic task characteristics and the process of WP-solving. For this we used eye-movement (fixation and regression time spent on the whole problem and on the text, numbers and keyword) in order to understand its direct influence as well as the possible common processing phases. The direct influence of difficulty could be conceptually replicated: The difficulty of WPs was increased by arithmetic factors in terms of the arithmetic operation, as well as by linguistic factors in terms of lexical consistency and nominalization. The results speak to an interaction between the initial reading process and the other problem-solving processes. Furthermore, task characteristics elicited an attentional shift in distinct eye-movement patterns targeting specific parts of the WP. Finally, we found interactions in the eye-movement behaviour: arithmetic complexity alone led to an increased eye-movement pattern on the linguistic part of the problem. This indicates that arithmetic processes are not fully separable form other problem-solving processes. How specific task characteristics influence problem-solving and how they are related to problem-solving processes will be further discussed in the following section.

INCREASED DIFFICULTY OF WORD PROBLEMS: BEHAVIORAL RESULTS

Behavioural studies have shown that both linguistic and arithmetic task characteristics contribute (directly or via interaction) to the difficulty of a WP. Our results replicated these findings.

Operation effect. We found a general increase in reaction time for subtraction as compared to addition. This is in line with the finding that subtraction elicits greater response times (Orrantia et al., 2012) than addition for arithmetic WPs. Although, subtraction is generally more difficult (Artemenko et al., 2015) in this study, in terms of ACC, the effects of operation were found only for sentences with the nominalized form (i.e., semantically more complex sentences) but not for sentences with the verbalised form. This can be explained by the fact that some types of simple WP language modification, without changing the arithmetic structure, might lead to higher success rates (Cummins et al., 1988; Davis-Dorsey et al., 1991; Vicente et al., 2007). However, these studies did not investigate the general influence of rewording on the arithmetic operation and our results suggest that the complexity of operation can be influenced by a simple rewording of the text.

Nominalization effect. Regarding the linguistic difficulty of the WPs, we found a general increase in the reaction time for nominalization. This extends previous findings of other studies that found nominalizations to be a source of difficulty (Abedi et al., 2005; Prediger et al., 2015).

Additionally, contrary to studies that found effects of nominalization in complex multi-step WPs only with a higher density of nominalization as compared to lower density of nominalization (Schlager et al., 2017), our study shows that even small changes can lead to a difference in the solution process, suggesting that simply rewording a problem into a nominalized form influences problem complexity.

For ACC, there was no main effect of nominalization, but we found interactions with consistency and operation only in nominalized sentences. This shows that when the text is easy to read and comprehend children are able to solve WPs which require a more complex mental representation. However, when the text has the same underlying arithmetic structure but is more difficult to read, the additional difficulty of consistency and operation increase the probability of errors (see also Vicente et al., 2008). suggests that in these cases it gets more difficult to create an appropriate mental representation. This is also supported by the finding that even inept problem solvers are able to solve difficult WPs when the text is semantically easy (Van der Schoot et al., 2009).

Consistency effect. Regarding the linguistic difficulty of the WPs, we found a general increase in the reaction time for nominalization, but not for lexical consistency. This result is contrary to some previous studies (Hegarty et al., 1992; Hegarty et al., 1995; Pape, 2003; Van der Schoot et al., 2009; Verschaffel et al., 1992), where WPs of inconsistent form elicited longer reaction times as compared to problems of consistent form, but in line with the study of Verschaffel et al. (1992) who found no consistency effect in adults. However, we found the consistency effect only for nominalized WPs, suggesting, as described above, that the presence of the consistency effect might depend on the formulation of the text, so that the difficulty of creating a problem model increase. This might also explain why some studies find the consistency effect (e.g., Van der Schoot et al., 2009) and others do not (e.g., Verschaffel et al., 1992).

In sum, our behavioural results show a direct influence (but no interactions) in the case of response time. On the other hand, the results on ACC show that common processing stages are possible: we found effects of operation and consistency only in the case of the semantically more complex, nominalized sentences. The behavioural results also speak to an overlap of the initial reading phase and the other phases of problem-solving. Namely, for ACC the effects of consistency and operation were found only for nominalized sentences, but not for verbalized sentences, suggesting a common processing stage. As nominalization could be associated more with the initial reading phase this means that in case of children it is hard to clearly distinguish between the initial reading process, especially when the reading demand increases, and other problem-solving processes. This indicates that the difficulty of the text formulation influences the creation of the problem model, i.e., a simple rewording can make it difficult to find the correct arithmetic mental model. The interaction between operation and nominalization indicates that in the case of children arithmetic processes interact with the initial reading processes. However, behavioural studies do not reflect which parts of the text play an important part in creating the arithmetic mental model and even in the case of an interaction we do not know whether it comes from arithmetic or linguistic difficulties.

INCREASED DIFFICULTY OF WORD PROBLEMS: GENERAL EYE-MOVEMENT

In this study we aimed to replicate the direct influence of textual difficulty on WP-solving performance (Hegarty et al., 1995; Strohmaier et al., 2019). Additionally, we hypothesised that task characteristics should elicit an attentional shift in the corresponding distinct eye-movement patterns targeting specific parts of the WP in general.

Eye-movement pattern over the whole problem. It has been already shown that numerous and difficult stimuli in mathematics lead to longer fixations and regression times (Knoblich et al., 2001). For WPs in particular, there is a relationship between increased task difficulty and increases in eye-movement measures related to the whole text (Strohmaier et al., 2019). Here, we found similar results, namely that subtraction problems elicited a longer total fixation time as well as longer regression on the whole problem than addition problems. This is in line with De Corte et al. (1990), who found that students spent more fixation time on the whole WP when it involved subtraction as compared to addition. As regressions are hypothesised to represent additional processing, this result suggests that subtraction requires extra processing and that the calculation process is probably extensively present. Additionally, we also showed that nominalized sentences elicit more total fixation time on the whole problem compared to verbalization. This demonstrates that even small textual changes influence the eye-movement behaviour. Finally, in the conclusion of the study of Strohmaier et al. (2019) higher mean fixation time was interpreted as reflecting either the direct translation strategy or struggling with the WP, and relatively short mean fixation as reflecting the problem model strategy. In our analysis only, the correct items were included, so a correct mental representation can be assumed. Therefore, the observation that linguistically and arithmetically more complex items elicited longer mean fixation time show that the item characteristics of WPs play an essential role in the eye-movement behaviour.

Arithmetic difficulty and the mathematical part of the problem. Despite the major role of arithmetic properties in mathematical cognition, arithmetic difficulty has not been investigated extensively for WP tasks with eye-movement (Fuchs et al., 2010). Nevertheless, we found that arithmetic difficulty increased the eye-movements towards the numbers in the WP in terms of both total fixation time and regression time. Namely, subtraction elicited more fixation time on the numbers than addition. This is in line with De Corte and Verschaffel (1986) who observed a relatively strong focus on the numbers in WPs.

Linguistic difficulty and the textual part of the problem. Linguistic difficulty increased fixation time and regression time on the text in the WP. For instance, the linguistic factor nominalization elicited greater attention towards the whole text without numbers in the case of total fixation time but not in the case of regression. Eye-movements towards the keyword showed a consistency effect, only in case of nominalization, on both fixation and regression time. Additionally, the consistency effect was found only for addition and an inverted consistency effect was found for subtraction problems. This finding is partly supported by previous research (Pape, 2003; Van der Schoot et al., 2009; Verschaffel et al., 1992) reporting a consistency effect only for addition and not subtraction. This shows that even a slight increase in the text difficulty (without changing the semantic congruency of the problem) elicits changes in the eye-movement behaviour. Interestingly, we did not observe the consistency effect in terms of total fixation on the whole WP but only on a specific textual part: the keyword. This indicates that the consistency effect does not manifest itself in the whole problem, but rather on a specific part of the problem and a specific phase of the solution process. This is in line with many studies that observed the consistency effect only for the second sentence containing the keyword (De Corte et al., 1990) or only in the second phase of problem-solving (e.g., Hegarty et al., 1992). Additionally, we found an interaction of nominalization in the fixation time on the overall problem and with consistency also in the regression time on the keyword. This means that the factor nominalization did not only influence the initial reading phase, but also that the comprehension of WPs possibly depends on the wording. This indicates that the process of reading is not completely separable from the process of solving and text formulation affects not only the first reading but also the other phases of the solution process, which is supported by the cyclic model (e.g., Bergqvist & Österholm, 2010).

In sum, WPs are not generally more difficult compared to other mathematical tasks but differ in the linguistic and arithmetic difficulty: Increasing arithmetic difficulty shifts eye-movements towards numerical elements and increasing linguistic difficulty shifts eye-movements towards the textual elements of WPs.

LINGUISTIC AND ARITHMETIC FACTORS INTERACT IN WORD PROBLEM-SOLVING

In this study we hypothesised that if the cognitive processes of reading and calculation are fully separable in terms of WPs, increased linguistic difficulty should lead to increased cognitive effort and attention on only the textual part of the problem (as indicated by eye-movement data) and mathematical processes should lead to increased cognitive effort and attention on only the numerical part of the problem. This was only partially true: linguistic difficulty did not affect the numerical areas, but arithmetic difficulty did affect the textual areas.

Linguistic difficulty and numerical areas. Contrary to what was expected, increased linguistic difficulty (lexical consistency or nominalization) did not increase the attention paid to the numbers. One explanation could be that lexical consistency affects other specific parts of the text, focusing more on the keyword. Another explanation could be that it depends on the language proficiency of the problem solver. For example, for non-native speakers linguistic difficulty has a significant influence on student perception of the arithmetic difficulty of problems (Barbu & Beal, 2010).

Arithmetic difficulty and textual areas. As the number of areas of interest were affected only by the factor operation this suggests at first glance that the arithmetic processing might not interfere with the other solution phases as the number of areas of interests were affected only by the factor operation. Nevertheless, for arithmetically more complex problems, an increase in the duration of fixation and regression on the textual elements was observed, which means that with increasing arithmetic difficulty the children looked at the text more. This can be interpreted to mean that arithmetic processes might also influence another process of problem-solving and subtraction makes it more difficult to create a representation of a WP and therefore people struggle not only with the numbers but also with the text. Bergqvist and Österholm (2010) also said that carrying out a calculation adds to the mental representation but does not change it, and they associated re-reading with changes to the existing mental representation. This also supports the idea that arithmetic processes do not only add to the mental model but change it. The explanation may be, that since subtraction is more difficult than addition (Artemenko et al., 2018) and there are less readily available strategies than in the case of addition, a larger portion of the cognitive resources will be gradually occupied as WPs work as a dual task, where the participant needs to simultaneously process both mathematics and language. If they allocate more resources for one task (math), there are less resources for the other (language). Therefore, there are probably less resources available to keep the non-arithmetic text in mind and the participant needs to spend a longer amount of time reading/looking at the text, while keeping and manipulating the subtraction problem in their mind.

LIMITATIONS & FUTURE DIRECTIONS

It is important to point out that in eye-movement research there is no ideal way of choosing a correct measure. For example, eye-movement studies usually use word level analysis, which is a controversial measure (Rayner, 1998). For instance, De Corte and Verschaffel (1986) showed that many people did not even look at important areas of the text problems and yet they answered correctly. This means that word level analysis does not necessarily contain all of the information about the processing. Many words are skipped, so the definition of fixation on all the words is not always enough. On the other hand, gaze duration (i.e., the sum of all fixation time) contains not only the reading but also the text integration processes. Last but not least we should be careful with the interpretation of fixation and regression times as well. On the one hand, the difficulty of the items generally increases fixation and regression times, but on the other hand it has also been shown that people with more expertise in a specific subject needed fewer eye fixations and regressions to process information in specific areas. Therefore, it would be preferable to investigate other eye-movement behaviours, such as the sequence of reading.

CONCLUSION

We conclude that eye-movement behaviour depends on the linguistic and arithmetic task characteristics. Complex and simple WPs generally result in different strategies, which can be detected in eye-movements. For instance, more complex problems reinforce longer fixations and regression times and are associated with lower performance. Even for children, WPs are not generally more difficult, but differ in the linguistic and arithmetic difficulty: Increasing arithmetic difficulty shifts eye-movements towards numerical elements and increasing linguistic difficulty shifts eye-movements towards the textual elements of a WP. However, the influence of linguistic and arithmetic factors is not reduced to main effects as interactions can be observed in specific parts of WPs. For instance, in arithmetically complex problems (e.g., with subtraction) children spend more time on the textual parts of the WP. However, how long children look at the numbers in the WP does not seem to be affected by linguistic difficulty.

Acknowledgments

This research was funded by the LEAD Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. We would also like to thank our research assistants Johanna Zeibig, Hanna Lönneker, Zoe-Lauren Kirste for their help in data collection and language proofreading of the manuscript.

SUPPLEMENTARY MATERIAL 1: WORD PROBLEMS

Table 3S1 - 1 *Stimuli*.

Block	Nr.	Sentences	Condition
B1			
	1	S1: Ein Mann spart Geld für einige Anschaffungen. *Translation: A man saves money on some purchases.* S2: Er hatte 23 Euro. *Translation: He had 23 Euros.* S3: Er verdient 34 Euro. *Translation: He earns 34 Euros.* S4: Wie viel Geld hat der Mann jetzt? *Translation: How much money does the man have now?*	V C A nC
	2	S1: Eine Studentin muss Wörter lernen. *Translation: A student must learn words.* S2: Die Studentin kannte 64 Wörter. *Translation: The student knew 64 words.* S3: Sie hat 27 Wörter gerade gelernt. *Translation: She just learned 27 words.* S4: Wie viele Wörter kennt die Studentin am Ende? *Translation: How many words does the student know in the end?*	V C A Ca
	3	S1: Eine Tennisspielerin spielt Turniere im Einzel. *Translation: A tennis player plays tournaments in singles.* S2: Sie hat in dieser Saison 79 Spiele gespielt. *Translation: She has played 79 games this season.* S3: Sie verlor 17 Spiele. *Translation: She lost 17 games.* S4: Wie viele Spiele hat sie gewonnen?	V C S nC
	4	S1: Ein Dieb hat einer Frau einige Diamanten gestohlen. *Translation: A thief stole some diamonds from a woman.* S2: Der Dieb hat ihr 14 Diamanten gestohlen. *Translation: The thief stole 14 diamonds from her.* S3: Sie hatte am Anfang 42 Diamanten gehabt. *Translation: She did have 42 diamonds in the beginning.* S4: Wie viele Diamanten hat sie jetzt?	V C S Ca

Translation: How many diamonds does she have now?

5 S1: Ein paar Freunde suchen Pilze im Wald. V I A nC

Translation: A few friends are looking for mushrooms in the forest.

S2: Sie werfen gerade 14 giftige Pilze weg.

Translation: They now throw away 14 poisonous mushrooms.

S3: Sie haben 73 Pilze.

Translation: They have 73 mushrooms.

S4: Wie viele Pilze hatten sie früher?

Translation: How many mushrooms did they have earlier?

6 S1: Eine Tennisspielerin spielt Turniere im Einzel. V I A Ca

Translation: A tennis player plays tournaments in singles.

S2: Sie hat in dieser Saison 18 Spiele verloren.

Translation: She has lost 18 games this season.

S3: Sie hatte früher 53 Spiele verloren.

Translation: She had previously lost 53 games.

S4: Wie viele Spiele hat sie in ihrer Karriere verloren?

Translation: How many games has she lost in her career?

7 S1: Ein Dieb hat einer Frau einige Diamanten gestohlen. V I S nC

Translation: A thief stole some diamonds from a woman.

S2: Der Dieb hatte ihr 58 Diamanten gestohlen.

Translation: The thief had stolen 58 diamonds from her.

S3: Die Polizei hat 26 Diamanten wiedergefunden.

Translation: The police recovered 26 diamonds.

S4: Wie viele Diamanten muss die Polizei noch finden?

Translation: How many diamonds are left for the police to find?

8 S1: Einige Leute wurden zur Party eingeladen. V I S Ca

Translation: Some people were invited to the party.

S2: 15 Leute sind gerade zur Party gekommen.

Translation: 15 people just came to the party.

S3: Am Ende sind 61 Leute auf der Party.

Translation: In the end there are 61 people at the party.

S4: Wie viele Leute waren am Anfang auf der Party?

Translation: How many people were in the beginning at the party?

9 S1: Ein Mädchen hat ihren Freunden Bücher mitgebracht. N C A nC

Translation: A girl brought her friends books.

S2: Das Mädchen hat 52 Bücher.

Translation: The girl has 52 books.

S3: Sie bedankt sich für das Erhalten von 16 Büchern.

*Glossing: She thanks for **the receiving** of 16 books.*

Translation: She thanks for receiving 16 books.

S4: Wie viele Bücher hat sie am Ende?

Translation: How many books does she have in the end?

10 S1: Ein paar Freunde suchen Pilze im Wald. N C A Ca

Translation: A few friends are looking for mushrooms in the forest.

S2: Sie hatten 14 Pilze.

Translation: They had 14 mushrooms.

S3: Sie sind gerade froh über das Sammeln von 58 Pilzen.

*Glossing: They are now happy about **the collection** of 58 mushrooms.*

Translation: They are now happy about collecting 58 mushrooms.

S4: Wie viele Pilze haben sie am Ende?

Translation: How many mushrooms do they have in the end?

11 S1: Eine Studentin muss Wörter lernen. N C S nC

Translation: A student must learn words.

S2: Die Studentin ist unglücklich über das Vergessen von 25 Wörtern.

*Glossing: The student is unhappy about **the forgetting** of 25 words.*

Translation: The student is unhappy about forgetting 25 words.

S3: Sie hat am Anfang 46 Wörter gelernt.

Translation: She learned 46 words in the beginning.

S4: Wie viele Wörter hat sie richtig gelernt?

Translation: How many words did she learn correctly?

12 S1: Eine Marktfrau verkauft Äpfel auf dem Markt N C S Ca

Translation: A market woman sells apples in the marketplace.

S2: Die Marktfrau kam mit 63 Äpfeln.

Translation: The market woman came with 63 apples.

S3: Sie freut sich über den Verkauf von 26 Äpfeln.

Translation: She is happy about the sale of 26 apples.

S4: Wie viele Äpfel hat sie am Ende?

Translation: How many apples does she have in the end?

13 S1: Eine Marktfrau verkauft Äpfel auf dem Markt N I A nC

Translation: A market woman sells apples in the marketplace.

S2: Die Marktfrau freut sich über den Verkauf von 23 Äpfeln.

Translation: She is happy about the sale of 23 apples.

S3: Sie hat danach 61 Äpfel.

Translation: She has 61 apples afterwards.

S4: Wie viele Äpfel hatte sie am Anfang?

Translation: How many apples did she have in the beginning?

14 S1: Ein Mann spart Geld für einige Anschaffungen. N I A Ca

Translation: A man saves money on some purchases.

S2: Er bedauert das Ausgeben von 17 Euro.

*Glossing: He regrets **the spending** of 17 Euros.*

Translation: He regrets spending 17 Euros.

S3: Er hat jetzt 45 Euro.

Translation: He now has 45 Euros.

S4: Wie viel Geld hatte der Mann?

Translation: How much money did the man have?

15 S1: Einige Leute wurden zur Party eingeladen. N I S nC

Translation: Some people were invited to the party.

S2: Es sind 39 Leute auf der Party.

Translation: There are 39 people at the party.

S3: 14 Leute entschieden sich für ein späteres Kommen.

*Glossing: 14 people decided for a later **coming**.*

Translation: 14 people decided to come later.

S4: Wie viele Leute waren am Anfang auf der Party?

Translation: How many people were in the beginning at the party?

16 S1: Ein Mädchen hat ihren Freunden Bücher mitgebracht. N I S Ca

Translation: A girl brought her friends books.

S2: Das Mädchen bedankt sich für das Erhalten von 27 Büchern.

*Glossing: The girl thanks for **the receiving** of 27 books.*

Translation: The girl thanks for receiving 27 books.

S3: Sie hat danach 56 Bücher.

Translation: She has 56 books afterwards.

S4: Wie viele Bücher hatte sie am Anfang?

Translation: How many books did she have in the beginning?

B2

17 S1: Ein paar Freunde suchen Pilze im Wald. V C A nC

Translation: A few friends are looking for mushrooms in the forest.

S2: Sie hatten 71 Pilze.

Translation: They had 71 mushrooms.

S3: Sie sammeln gerade 26 Pilze.

Translation: They are currently collecting 26 mushrooms.

S4: Wie viele Pilze haben sie am Ende?

Translation: How many mushrooms do they have in the end?

18 S1: Einige Leute wurden zur Party eingeladen. V C A Ca

Translation: Some people were invited to the party.

S2: Es waren anfangs 18 Leute auf der Party.

Translation: There were 18 people at the beginning of the party.

S3: 79 Leute sind später gekommen.

Translation: 79 people came later.

S4: Wie viele Leute waren auf der Party am Ende?

Translation: How many people were at the party in the end?

19 S1: Ein Mann spart Geld für einige Anschaffungen. V C S nC

Translation: A man saves money on some purchases.

S2: Er hat 12 Euro ausgegeben.

Translation: He spent 12 Euros.

S3: Er hatte 74 Euro gehabt.

Translation: He had 74 Euros.

S4: Wie viel Geld hat er am Ende?

Translation: How much money does he have in the end?

20 S1: Eine Marktfrau verkauft Äpfel auf dem Markt V C S Ca

Translation: A market woman sells apples in the marketplace.

S2: Die Marktfrau verkauft 29 Äpfel.

Translation: The market woman sells 29 apples.

S3: Sie hatte 65 Äpfel mitgebracht.

Translation: She brought 65 apples.

S4: Wie viele Äpfel hat sie am Ende?

Translation: How many apples does she have in the end?

21 S1: Ein Dieb hat einer Frau einige Diamanten gestohlen. V I A nC

Translation: A thief stole some diamonds from a woman.

S2: Der Dieb hatte ihr 36 Diamanten gestohlen.

Translation: The thief stole 36 diamonds from her.

S3: Sie hat jetzt 12 Diamanten.

Translation: She now has 12 diamonds.

S4: Wie viele Diamanten hatte die Frau?

Translation: How many diamonds did the woman have?

22 S1: Ein Mädchen hat ihren Freunden Bücher mitgebracht. V I A Ca

Translation: A girl brought her friends books.

S2: Das Mädchen verschenkt 26 Bücher.

Translation: The girl gifts 26 books.

S3: Sie hat danach 38 Bücher.

Translation: She has 38 books afterwards.

S4: Wie viele Bücher hatte sie am Anfang?

Translation: How many books did she have in the beginning?

23 S1: Ein Mann spart Geld für einige Anschaffungen. V I S nC

Translation: A man saves money on some purchases.

S2: Er hat jetzt 47 Euro.

Translation: He now has 47 Euros.

S3: Er hatte 15 Euro verdient.

Translation: He had earned 15 Euros.

S4: Wie viel Geld hatte er am Anfang?

Translation: How much money did he have in the beginning?

24 S1: Eine Studentin muss Wörter lernen. V I S Ca

Translation: A student must learn words.

S2: Die Studentin hat 14 Wörter gelernt.

Translation: The student learned 14 words.

S3: Sie kennt jetzt 82 Wörter.

Translation: She now knows 82 words.

S4: Wie viele Wörter hatte sie am Anfang gekonnt?

Translation: How many words did she know in the beginning?

25 S1: Eine Studentin muss Wörter lernen. N C A nC

Translation: A student must learn words.

S2: Die Studentin kannte 62 Wörter.

Translation: The student knew 62 words.

S3: Sie hat gerade das Lernen von 21 Wörtern gemeistert.

*Glossing: She mastered **the learning** of 21 words.*

Translation: She mastered learning 21 words.

S4: Wie viele Wörter kennt die Studentin am Ende?

Translation: How many words does the student know in the end?

26 S1: Eine Tennisspielerin spielt Turniere im Einzel. N C A Ca

Translation: A tennis player plays tournaments in singles.

S2: Sie hat vor dieser Saison 36 Spiele gewonnen.

Translation: She has won 36 games before this season.

S3: Sie freut sich über das Gewinnen von 17 Spielen in dieser Saison.

*Glossing: She is happy about **the winning** of 17 games this season.*

Translation: She is happy about winning 17 games this season.

S4: Wie viele Spiele hat sie in ihrer Karriere gewonnen?

Translation: How many games has she won in her career?

27 S1: Eine Tennisspielerin spielt Turniere im Einzel. N C S nC

Translation: A tennis player plays tournaments in singles.

S2: Sie spricht nicht gerne über das Verlieren von 32 Spielen.

*Glossing: She does not like to talk about **the loss** of 32 games.*

Translation: She does not like to talk about losing 32 games.

S3: Sie hat in dieser Saison 45 Spiele gespielt.

Translation: She has played 45 games this season.

S4: Wie viele Spiele hat sie gewonnen?

Translation: How many games did she win?

28 S1: Ein Mädchen hat ihren Freunden Bücher mitgebracht. N C S Ca

Translation: A girl brought her friends books.

S2: Das Mädchen beschließt das Verschenken von 12 Büchern.

*Glossing: The girl decides **the gifting** of 12 books.*

Translation: The girl decides to gift 12 books.

S3: Sie hatte 41 Bücher mitgebracht.

Translation: She brought 41 books.

S4: Wie viele Bücher hat sie am Ende?

Translation: How many books does she have in the end?

29 S1: Ein paar Freunde suchen Pilze im Wald. N I A nC

Translation: A few friends are looking for mushrooms in the forest.

S2: Sie entschlossen sich gerade für das Wegwerfen von 17 giftigen Pilzen.

*Glossing: They now decided for **the throwing** away of 17 poisonous mushrooms.*

Translation: They now decided to throw away 17 poisonous mushrooms.

S3: Sie haben 51 Pilze.

Translation: They have 51 mushrooms.

S4: Wie viele Pilze hatten sie davor?

Translation: How many mushrooms did they have earlier?

30 S1: Einige Leute wurden zur Party eingeladen. N I A Ca

Translation: Some people were invited to the party.

S2: 67 Leute entschieden sich für ein Verlassen der Party.

*Glossing: 67 people decided for **a leaving** of the party.*

Translation: 67 people decided to leave the party.

S3: Es blieben danach 15 Leute.

Translation: There remained 15 people afterwards.

S4: Wie viele Leute waren zur Party gekommen?

Translation: How many people had come to the party?

31 S1: Ein Dieb hat einer Frau einige Diamanten gestohlen. N I S nC

Translation: A thief stole some diamonds from a woman.

S2: Der Dieb hatte ihr 56 Diamanten gestohlen.

Translation: The thief had stolen 56 diamonds from her.

S3: Die Polizei feiert das Wiederfinden von 21 Diamanten.

Translation: The police celebrate the recovery of 21 diamonds.

S4: Wie viele Diamanten muss die Polizei noch finden?

Translation: How many diamonds are left for the police to find?

32 S1: Eine Marktfrau verkauft Äpfel auf dem Markt N I S Ca

Translation: A market woman sells apples in the marketplace.

S2: Die Marktfrau hat 34 Äpfel.

Translation: The market woman has 34 apples.

S3: Die Marktfrau hatte sich vorher über die Lieferung von 19 Äpfeln gefreut.

Translation: The market woman had been looking forward to the delivery of 19 apples before.

S4: Wie viele Äpfel hatte sie am Anfang?

Translation: How many apples did she have in the beginning?

B3

33 S1: Ein Dieb hat einer Frau einige Diamanten gestohlen. V C A nC

Translation: A thief stole some diamonds from a woman.

S2: Nach dem Diebstahl hatte sie 31 Diamanten.

Translation: After the theft she had 31 diamonds.

S3: Die Polizei findet 17 Diamanten wieder.

Translation: The police recover 17 diamonds.

S4: Wie viele Diamanten hat die Frau am Ende?

Translation: How many diamonds does the woman have in the end?

34 S1: Ein Mädchen hat ihren Freunden Bücher mitgebracht. V C A Ca

Translation: A girl brought her friends books.

S2: Das Mädchen hat 34 Bücher.

Translation: The girl has 34 books.

S3: Sie erhält 19 Bücher.

Translation: She receives 19 books.

S4: Wie viele Bücher hat sie am Ende?

Translation: How many books does she have in the end?

35 S1: Ein Mädchen hat ihren Freunden Bücher mitgebracht. V C S nC

Translation: A girl brought her friends books.

S2: Das Mädchen hat 76 Bücher.

Translation: The girl has 76 books.

S3: Sie verschenkt 15 Bücher an ihre Freunde.

Translation: She gifts 15 books to her friends.

S4: Wie viele Bücher hat sie am Ende?

Translation: How many books does she have in the end?

36 S1: Einige Leute wurden zur Party eingeladen. V C S Ca

Translation: Some people were invited to the party.

S2: 56 Leute sind zur Party gekommen.

Translation: 56 people came to the party.

S3: 19 Leute haben die Party früh verlassen.

Translation: 19 people left the party early.

S4: Wie viele Leute sind jetzt auf der Party?

Translation: How many people are at the party now?

37 S1: Eine Marktfrau verkauft Äpfel auf dem Markt V I A nC

Translation: A market woman sells apples in the marketplace.

S2: Die Marktfrau verkauft 12 Äpfel.

Translation: The market woman sells 12 apples.

S3: Sie hat danach 74 Äpfel.

Translation: She has 74 apples afterwards.

S4: Wie viele Äpfel hatte sie am Anfang?

Translation: How many apples did she have in the beginning?

38 S1: Ein Mann spart Geld für einige Anschaffungen. V I A Ca

Translation: A man saves money on some purchases.

S2: Er hat 16 Euro ausgegeben.

Translation: He spent 16 Euros.

S3: Er hat jetzt 59 Euro.

Translation: He now has 59 Euros.

S4: Wie viel Geld hatte der Mann?

Translation: How much money did the man have?

39 S1: Eine Tennisspielerin spielt Turniere im Einzel. V I S nC

Translation: A tennis player plays tournaments in singles.

S2: Sie hat in dieser Saison 39 Spiele gespielt.

Translation: She has played 39 games this season.

S3: Sie gewann 21 Spiele.

Translation: She won 21 games.

S4: Wie viele Spiele hatte sie verloren?

Translation: How many games had she lost?

40 S1: Ein paar Freunde suchen Pilze im Wald. V I S Ca

Translation: A few friends are looking for mushrooms in the forest.

S2: Sie haben gerade 28 Pilze gesammelt.

Translation: They just collected 28 mushrooms.

S3: Sie haben am Ende 52 Pilze.

Translation: They have 52 mushrooms in the end.

S4: Wie viele Pilze hatten sie davor?

Translation: How many mushrooms did they have before?

41 S1: Einige Leute wurden zur Party eingeladen. N C A nC

Translation: Some people were invited to the party.

S2: Es waren anfangs 26 Leute auf der Party.

Translation: There were 26 people at the beginning of the party.

S3: 43 Leute haben sich für ein späteres Kommen entschieden.

Glossing: 43 people have decided for ***a later coming****.*

Translation: 43 people have decided to come later.

S4: Wie viele Leute waren auf der Party am Ende?

Translation: How many people were at the party in the end?

42 S1: Eine Marktfrau verkauft Äpfel auf dem Markt N C A Ca

Translation: A market woman sells apples in the marketplace.

S2: Die Marktfrau kam mit 28 Äpfeln.

Translation: The market woman came with 18 apples.

S3: Sie freut sich über eine Lieferung von 64 Äpfeln.

Translation: She is happy about a delivery of 64 apples.

S4: Wie viele Äpfel hat sie am Ende?

Translation: How many apples does she have in the end?

43 S1: Ein Dieb hat einer Frau einige Diamanten gestohlen. N C S nC

Translation: A thief stole some diamonds from a woman.

S2: Die Frau hatte 37 Diamanten gehabt.

Translation: The woman has had 37 diamonds.

S3: Sie war über das Stehlen von 12 Diamanten entsetzt.

Translation: She was horrified over the theft of 12 diamonds.

S4: Wie viele Diamanten hat sie jetzt?

Translation: How many diamonds does she have now?

44 S1: Ein paar Freunde suchen Pilze im Wald. N C S Ca

Translation: A few friends are looking for mushrooms in the forest.

S2: Sie haben 53 Pilze gefunden.

Translation: They found 53 mushrooms.

S3: Sie entschlossen sich für das Wegwerfen von 19 giftigen Pilzen.

Glossing: They decided for ***the throwing*** *away 19 poisonous mushrooms.*

Translation: They decided to throw away 19 poisonous mushrooms.

S4: Wie viele Pilze haben sie danach?

Translation: How many mushrooms do they have afterwards?

45 S1: Eine Tennisspielerin spielt Turniere im Einzel. N I A nC

Translation: A tennis player plays tournaments in singles.

S2: Sie spricht nicht gerne über das Verlieren von 56 Spielen in dieser Saison.

Glossing: She does not like to talk about ***the loss*** *of 56 games this season.*

Translation: She does not like to talk about losing 56 games this season.

S3: Sie hatte früher 23 Spiele verloren.

Translation: She had previously lost 23 games.

S4: Wie viele Spiele hat sie in ihrer Karriere verloren?

Translation: How many games has she lost in her career?

46 S1: Eine Studentin muss Wörter lernen. N I A Ca

Translation: A student must learn words.

S2: Die Studentin ist unglücklich über das Vergessen von 45 Wörtern.

Glossing: The student is unhappy about ***the forgetting*** *of 45 words.*

Translation: The student is unhappy about forgetting 45 words.

S3: Sie lernt später 17 andere Wörter.

Translation: She later learns 17 other words.

S4: Wie viele Wörter müsste sie kennen?

Translation: How many words should she know?

47 S1: Eine Studentin muss Wörter lernen. N I S nC

Translation: A student must learn words.

S2: Die Studentin hat das Lernen von 32 Wörtern gemeistert.

Glossing: The student mastered ***the learning*** *of 32 words.*

Translation: The student mastered learning 32 words.

S3: Sie kennt jetzt 46 Wörter.

Translation: She now knows 46 words.

S4: Wie viele Wörter hatte sie am Anfang gekonnt?

Translation: How many words did she know in the beginning?

48 S1: Ein Mann spart Geld für einige Anschaffungen. N I S Ca

Translation: A man saves money on some purchases.

S2: Er freut sich über das Verdienen von 16 Euro.

*Glossing: He is happy about **the earning** of 16 Euros.*

Translation: He is happy about earning 16 Euros.

S3: Er hat jetzt 35 Euro.

Translation: He now has 35 Euros.

S4: Wie viel Geld hatte er am Anfang?

Translation: How much money did he have in the beginning?

In the case of sentences with the nominalised form both the glossed and translated text are provided when necessary. S1: First sentence. S2: Second sentence. S3: Third sentence. S4: Fourth Sentence. V: Verbalised Form. N: Nominalised Form. C: Consistent Form. I: Inconsistent Form. S: Subtraction. A: Addition. Ca: Carry/Borrow. nC: Non-Carry/Non-Borrow.

SUPPLEMENTARY MATERIAL 2: TASK CONSTRAINT

There is few information about how the task characteristics shapes these eye-movement patterns towards the specific areas of the text as well as how this differentially affects the keywords and the numbers. There was not separation in most of the previous studies between keywords and the numbers (i.e., Hegarty et al., 1992; Hegarty et al., 1995). Simply because carrying out the exact calculation was not part of the task. However, when calculation phase if involved, that might result in a different eye-movement pattern. The reason is that only the visual input determines eye-movement patterns but task constraint as well (Yarbus, 2013), which means that we can have different eye-tracking patterns if we include the calculation or not. Also, Verschaffel et al. (1992) has shown that in a key-word searching strategy, i.e., subjects do not apply mental models, but simply search for the keyword with which a partial arithmetic operation is associated. Therefore, we argue that the separation of the keyword and the numbers might be necessary, especially when in the task constraint the calculation is involved. Not last, but least we aim whether the keyword and the numbers could be treated as one interest area or should be separated.

RESULTS

For *fixation duration looking at both keyword and numbers* (I4), there was a main effect for operation, $F(1,33) = 16.92$, $p < .001$, $\eta_p^2 = .35$, indicating that subtraction (8999 ms) elicited longer fixations than addition (7074 ms).

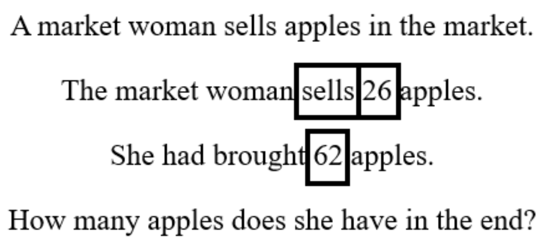

Figure 3S2 - 1 Combined Areas of Interest (CAOI): Numbers and Keyword together (I4).

For *regression time looking at both keyword and numerical interest* area (I4), there was a significant main effect for operation, $F(1,33) = 16.12$, $p < .001$, $\eta_p^2 = .34$, indicating that subtraction (7834 ms) elicited longer fixations than addition (5933 ms). Figure 1. Shows the interest areas keyword and numbers.

DISCUSSION

It is important to point out that in the studies investigating the formation of problem model strategies, no separation exists between numbers and keyword-words (considered together as one interest area). However, our results show a difference between the interest areas I3 (keyword word) and I4 (keyword word + number). For I4 contrary as expected, we do not see consistency effect, but for I3 we can observe. We also do not see the consistency effect on the numbers (I1), but we can observe for I3 the two-way interaction of the factor consistency in both in combination with nominalisation and operation. For I1 and I4 we see the same operation as a main effect. For I1 (numbers) and I4 (keyword) the results are similar. One explanation is that numbers (most probably due to the intensive demand on calculation) dominate the results of the I4 area. We can conclude that we should distinguish between studies that include the calculation phase or between studies where the task is only to state the planned calculation (Hegarty et al., 1992) because this may have an influence on the eye-tracking pattern. Therefore, we argue that the keyword and numerical area should be separately analysed when the calculation process is included.

SUPPLEMENTARY MATERIAL 3: PROBLEM-SOLVING

Eye-movement studies on WPs also investigated problem-solving strategies. For instance, it has been shown that high ability students differ in problem-solving strategies on the keywords and numbers; and other part of the text respectively (De Corte et al., 1990; Hegarty et al., 1992; Van der Schoot et al., 2009; Verschaffel et al., 1992).

Based on previous studies we would like to see whether more difficult problems change the regressions towards the keywords and numbers and the rest of the text (whole problem-keyword-number) respectively. These specific parts of WPs are associated with certain problem-solving strategies. For example, in the re-reading phase, the bias towards keywords and numbers is associated with direct translation or problem model strategies (Hegarty et al., 1992). Namely, students using direct translation would focus most heavily on the keywords and numbers in a problem rather than on the other words, while a student using the problem-model strategy would pay more attention to the other words in the problem in the difficult condition (Hegarty et al., 1992) when they re-read the WPs (Hegarty et al., 1995). Additionally, Van der Schoot et al. (2009) showed that less successful problem solvers make longer and more regressions in inconsistent than consistent WPs, but only when the problems contained unmarked relational terms. Because problems that are more difficult need higher text comprehension skills (especially in case of consistency, eventually nominalisation) there should be generally more regression on the above-mentioned textual areas. When the assumptions above are true and the bias towards these specific areas reflect problem-solving strategies, we expect more regression on the rest of the text (in this case all text minus the cue word and number) but not on the numbers and cue word in more difficult conditions, especially those which require higher text comprehension skills (e.g. consistency, nominalization).

RESULTS

For *regression time looking at all the textual elements without numbers and cue* (I5), there were significant main effects for nominalization $F(1,33) = 5.01$, $p = .002$, $\eta_p^2 = .26$, indicating that nominalisation (7755 ms) leads to longer regressions than verbalisation (7026 ms), and for operation $F(1,33) = 11.06$, $p = .032$, $\eta_p^2 = .14$, indicating that addition (7929 ms) leads to longer fixations than compared to subtraction (6852 ms).

For the interest areas numbers (I1), keyword (I2) the results can be found in the main article, the results for numbers and keyword (I4) in the Supplementary Material 2.

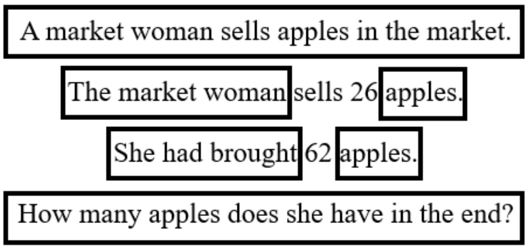

Figure 3S3 - 1 Combined Areas of Interest (CAOI): Textual information without Numbers and Keyword (I5).

DISCUSSION

We expected more regressions on the “text minus the cue and number” (I5) areas in case of the more difficult problems compared to the less difficult counterpart. This could be partially confirmed. Subtraction sentences elicited longer regressions than in sentences with addition. Also, nominalised sentences resulted in longer regressions times than verbalised cases. According to Hegarty et al. (1992), this could mean, that in the more difficult cases generally increase the comprehension need. Not only the linguistics, semantic factor, but also the mathematical difficulty. For consistency, no such results could be found.

SUPPLEMENTARY MATERIAL 4: INDIVIDUAL CHARACTERISTICS AND EYE-MOVEMENT PATTERNS

It is assumed that internal cognitive processing is synchronised with eye fixations (Just & Carpenter, 1987). Several studies on WPs have used eye-movement methodology to investigate underlying cognitive processing (De Corte et al., 1990; Hegarty et al., 1992; Verschaffel et al., 1992). These papers separate the initial reading phase from the other problem phases in order to locate where students experience difficulties (Kintsch & Greeno, 1985). Importantly, these different phases of WP-solving and the separation of the initial reading phase from other phases have been postulated but not tested empirically. Finally, it is also hypothesised that different individual abilities reinforce different reading patterns. Therefore, it is worth investigating whether successful reading patterns can be observed when the initial reading phase is not present.

Eye-movement studies on WPs have divided the problem-solving process into at least two parts: the first reading phase, and the following phases. This assumption is based on the idea that problem solvers first must read the problem before they can start solving it (Kintsch & Greeno, 1985). Most of the eye-movement studies on WPs have investigated where the consistency effect, based on the hypothesis of Lewis and Mayer (1987), originates from, and where to locate student difficulty (Hegarty et al., 1992; Hegarty et al., 1995; Verschaffel et al., 1992). It is important to point out, however, that for the separation of sequences of problem-solving with the help of eye-movement there is no theoretical consensus, and different definitions might lead to different results; therefore, further investigation is required. Most of the papers in the field of eye-movement and WP research support the dominance of systematic sequential problem-solving strategies. However, there is evidence that the first initial reading, or the whole problem-solving does not necessarily happen *sequentially* (Thevenot & Oakhill, 2006).

INDIVIDUAL CHARACTERISTICS AND EYE-MOVEMENT PATTERNS

No researchers have investigated how different eye-movement behaviour (i.e., non-sequential or sequential initial reading pattern) correlates to individual abilities and solution success. However, De Corte and Verschaffel (1986) also suggest a relationship between the problem-solving ability of children and their initial reading pattern. They found that most rereading came from children with better WP-solving skills. This would suggest that individual characteristics influence whether a systematic initial reading phase can be observed, and that individuals who solve WPs with high ACC more often use non-sequential reading patterns. However, in the eye-movement studies on WPs, differences between high and low ability students were partly based on task performance itself. In very few exceptions, stimulus independent tests were used. High and low ability are selected, for example in the Van der Schoot et al. (2009) or Hegarty et al. (1995), based on the WP-solving performance. However, the general mathematical abilities of the participants in the above-mentioned studies were not assessed, so little is known about the relation of arithmetic ability to the eye-movement pattern as regards WPs. Various other cognitive variables (e.g., intelligence, reading speed and visual and verbal working memory) necessary for problem-solving have not been investigated before in eye-movement studies on WPs. Since eye-movement behaviour is influenced for instance by reading comprehension skills, eye-movement behaviour and even simple reading strategies can differ. (Rayner, 1998). In sum, it is important to be able to discriminate between various phases of problem-solving, and to link individual characteristics to reading strategies.

In previous eye-movement studies several individual patterns were observed: WPs can be read sequentially or non-sequentially before arriving at the end of the first reading. An example can be found in De Corte et al. (1990): the pupil looked at the two numbers before starting to read the first sentence, and then reread the first two sentences before looking at the third one. People might also skip parts of the WPs. However, this skipping of information does not necessary lead to a lower solution rate. Salvucci (2001) showed that for algebra problems participants not only fixated in left-to-right patterns, but also sometimes skipped, and at other times re-fixated on, difficult items. This also applies to WPs: De Corte and Verschaffel (1986) showed that one child answered correctly even without reading the question. Additionally, the question plays a specific role in the solving of the WP (Cook & Rieser, 2005) and some people might start the reading phase by looking at the question first: highly capable students used question-guided strategies more often as compared to students with lower WP-solving ACC. This leads to our hypothesis: Starting with the question improves the solution rate if people have good working memory and reading abilities. In sum, we hypothesize that individual skills will be a strong positive predictor of the use of non-sequential strategies. Furthermore, it is assumed that successful and unsuccessful problem solvers will have different eye-movement patterns on different interest areas (Hegarty et al., 1995). There is evidence that non-sequential reading might lead to better solution strategies (Thevenot et al., 2007) because different people might use different solution strategies.

Concluding, we can say that studies about WP-solving often assume that participants use sequential reading strategies. Moreover, no study is known by the authors that investigated which children tend to use sequential reading strategies or non-sequential reading strategies, and how these strategies correlate to individual abilities. We hypothesise that there are individual differences in reading strategies and that these differences reinforce other reading patterns. Therefore, we conducted an exploratory analysis into how many participants start reading the question before reading other sentences in the first reading phase and whether this correlates with individual abilities.

DATA ANALYSIS

We used the sequence of fixation on the words and sentences to define the sequence of reading and to detect different reading patterns according to De Corte (1991). Here, an important decision was whether to consider the initial reading phase as happening in a strictly systematic sequential manner or whether participants might start with a sentence other than the first one, and if regressions to previously read sentences occurred or not. We categorized all reading patterns with regressions into non-sequential strategies. Only successfully solved items were considered for the analysis. We defined the following reading patterns:

- Sequential Reading Strategies
  - *Reading word by word:* Each word has to be fixated on at least once, and no regressions to previously read words are allowed.
  - *Reading sentence by sentence:* Each sentence has to be fixated on at least once, and no regressions to previously read sentences are allowed.
- Non-Sequential Reading Strategies
  - *Reading word by word but regressions are allowed*: Each word has to be fixated on at least once and regressions to previously read words are allowed.
  - *Reading sentence by sentence, but regressions are allowed.* Each word has to be fixated on least once, and regressions to previously read words are allowed.
  - *Question read before fixating on other sentences.* There is at least one fixation on the question before other sentences are fixated on.

RESULTS

Multiple linear regressions were calculated to predict the reading patterns based on the cognitive abilities.

Sequential Reading Strategies

1. Reading word by word. F(2,28)=7,179 p= .003, with an R^2 .339 adjusted R .292 Visual -.052 C 2,34 Beta -.355 Reading speed -.001 Beta -.363 Visual working memory and reading speed were negative significant predictors.

2. Reading sentence by sentence. F(1,29)=4,345 p= 0.046, with an R^2 0.131 adjusted R 0.100 LS-RÜCK -0.075 C 1,656 Beta -0.361. LS-Backword is associated with working memory.

Non-Sequential Reading Strategies

Reading word by word but regressions are allowed. F(2,28)=16,983 p= .000, R^2 .548 adjusted R .516 Subtraction skill-.034 C 2,55 Beta -.284 Reading speed -.001 Beta -.602 Subtraction skill and reading speed was a negative significant predictor.

Reading sentence by sentence, but regressions are allowed. F(1,29)=5,589 p= 0.025, R^2 0.162 adjusted R 0.133 Reading Comprehension -0.029 C 2,540 Beta -.361 Reading Comprehension was a negative significant predictor.

Question read before fixating on other sentences. F(1,29)=5,518, p= .026, with an R^2.160 (adjusted R .131). Reading comprehension was a positive significant predictor.

DISCUSSION

We investigated how individual abilities reinforce different reading patterns. We showed that individual capabilities predict different reading strategies; that is, reading comprehension predicted non-sequential strategies. The hypothesis that there are individual differences in reading strategies was confirmed. Reading speed, reading comprehension and working memory turned out to be the strongest predictors.

Sequential Reading Strategies. Working memory was a negative predictor in the reading strategy: Reading the sentences sequentially without returning to previously read sentences before fixating on the question. The role of working memory is supported by the study of Hegarty et al. (1995) in which the authors used a naming and remembering task. Hegarty et al. (1995) showed that when successful problem solvers make errors in remembering WPs, they are more likely than are unsuccessful problem solvers to remember the situation described in the problem but less likely than unsuccessful problem solvers to remember the specific relational keyword used in the problem (e.g. less or more). Reading comprehension and reading speed also played a part in our study: slower reading speed and lower reading comprehension skills were associated with more frequent use of the reading strategy.

Non-Sequential Reading Strategies. We suggested that reading the question before other sentences helps one to solve the problem correctly, especially if solvers have good working memory and reading abilities. A multiple linear regression confirmed that better reading comprehension was associated with more frequent use of reading patterns that were non-sequential. This is supported by (Thevenot et al., 2007) who observed better performance when the question was presented prior to the text especially in the case of children who had poorer mathematical skills and in the case of more difficult problems.

Additionally, we found that the variance of patterns is influenced by individual characteristics. The eye-movement analysis showed various reading and solution strategies such as sequential reading, non-sequential reading, or skipping the question, spending more time in between regressions of numbers and so on. This is in line with Cook and Rieser (2005), who found that high achieving students flexibly varied their visual scanning to fit problem difficulty, while low achievers showed less flexibility. This means that persons who achieve high solution ACC apply more strategies and show a broader variety of reading patterns. The non-sequential patterns in this study do not include all possible patterns; therefore, further investigation is necessary to explore other possible reading strategies. Our results are in line with De Corte et al. (1990) who suggested that we need to question the sequential and linear character of our theoretical model of competent problem-solving, especially with respect to more complex problem types. This points out that the separation of the first reading phase from the other problem-solving phases is not straightforward because information processing might also occur in the initial reading phase, so that the initial reading and problem-solving phases might overlap. This is also supported by Cook and Rieser (2005) where out of 28 students, 8 had already discriminated information in the initial reading phase. Most studies agree that the integration and planning phases overlap and are harder to isolate from the problem-solving process. However, it is important to note that there is no common methodology for differentiating reading time from the problem-solving process in reading research literature (Rayner, 1998). Although several studies agree (e.g., Hegarty et al., 1992; Jaffe, 2006) that it is possible to separate the phases of problem-solving via eye-movement, no study has investigated in detail whether this separation with eye-movement is actually possible and how various definitions of the phases affect the results. For instance, De Corte (1991) decided to include in their analysis only those eye-movement protocols that showed several consecutive fixations on the first sentence and on the second sentence before the eye moved to the third sentence. A considerable amount of eye-movement protocols did not meet this criterion, namely 80 out of 480. Also, in De Corte and Verschaffel (1986) 15 of 66 tasks were not fully read, and in De Corte et al. (1990) only those children who initially read through at least 10 of 16 target problems were included in the analysis; only half of the children met these criteria, and 201 out of the 320 protocols could be categorised into a pattern that contained a first reading phase. In total, 119 eye-movement protocols did not start with a systematic reading phase.

In conclusion, individual abilities influence eye-movement patterns, and further research is needed – with the help of more advanced techniques – to explore and categorise individual reading patterns. To understand how and under what circumstances people solve WPs, the integration of individual abilities, task characteristics and eye-tracking measures – eventually leading to reading models and specific reading characteristics influencing eye-movement during reading – is needed.

STUDY 4: THE RELATION OF ENVIRONMENTAL FACTORS TO THE TASK DIFFICULTY IN WORD PROBLEMS

CONTRIBUTIONS OF CO-AUTHORS AND OTHER PERSONS

Authors

Gabriella Daroczy[1,3], Benjamin Fauth[3,4,5], Krzysztof Cipora[1,3,6], Detmar Meurers[2,3] & Hans-Christoph Nürk[1,3]

[1]Department of Psychology, Eberhard Karls University Tuebingen, Tuebingen, Germany

[2]Department of Linguistics, Eberhard Karls University Tuebingen, Tuebingen, Germany

[3]LEAD Graduate School & Research Network, Eberhard Karls University Tuebingen, Tuebingen, Germany

[4] Institute for Educational Analysis (IBBW), Stuttgart, Germany

[5] Hector Research Institute of Education Sciences and Psychology, Eberhard Karls University Tuebingen, Tuebingen, Germany

[6] Centre for Mathematical Cognition, Loughborough University, Loughborough, UK

Author	Author position	Scientific idea (%)	Data generation (%)	Analysis and Interpretation (%)	Paper writing (%)
Daroczy, G.	1	55	100[4]	60	70
Fauth, B.	2	20	0	20	10
Meurers, W. D.	3	10	0	5	0
Cipora, K.	4	5	0	5	10
Nuerk, H.-C.	5	10	0	10	10

Data generated in the course of student work

Bettina Held-Ulrich, Bachelor-Thesis, Title: *„Individuelle, sprachliche und mathematische Einflüsse auf das Lösen von Textaufgaben“*, 2016 SS

Proof-Reading

Zoë-Lauren Kirste

Student Research Assistants

Hanna Lönneker, Bettina Held-Ulrich and Jacqueline Jaus helped in data collection.

I/We certify that the above-stated is correct.

28.09.2020

Date, Signature of the doctoral committee or at least of one of the supervisors

[4] Responsible, student research assistants and Bettina Held-Ulrich helped in data collection (50%).

ABSTRACT

An important part of any math curriculum is solving word problems. Such problems are particularly difficult for many students. It has been shown that the difficulty of word problems arises from the characteristics of the problems themselves (e.g., linguistic and arithmetic characteristics) and the solver. However, very little is known about how the environmental factor teaching quality (i.e., cognitive activation, supportive climate and classroom management) influences successful solution of word problems. Therefore, in a sample of 387 5th and 6th grade students we measured ability to solve word problems in such a way that we independently manipulated the linguistic (lexical consistency and nominalization) and arithmetic (operation and carry/borrow) complexity. Individual characteristics (intelligence, motivation, mathematical and reading skill, socio-economic status) was also measured. The results showed that both individual capability and environmental factors influenced word problem performance. Intelligence, reading and mathematical skills positively influenced the overall accuracy. In the face of nominalization, the teaching quality scales "cognitive activation" and "supportive climate" were associated with higher solution accuracy of problems with greater reading difficulty. These results provide evidence for the importance of environmental factors, and not only individual factors, on performance on linguistically demanding word problems.

Keywords

Arithmetic word problem, teaching quality, individual characteristics, task difficulty

INTRODUCTION

Solving word problems is an essential part of the educational goals of mathematics curricula across the world (Stacey, 2005) as a means to develop general problem-solving skills. Good problem-solving skills are a strong positive predictor of young adult ability to participate in post-secondary education (OECD, 2013) as well as the best predictor of adult employment. However, the goal of many curricula, to teach complex problem-solving abilities, cannot be met because students experience severe difficulty with simple arithmetic WPs. For instance, according to OECD (2013), 32% of students were not able to extract relevant information from a sole source to solve WPs involving whole numbers.

Although WPs are one of the most researched topics in the field of mathematics education, much is unknown about the origin of their difficulty and why some individuals are more successful in solving WPs than others. A crucial element is how the text is formulated. Both linguistic and arithmetic factors (i.e., task characteristics) influence the ACC of a solution; however, depending on individual characteristics, these factors are differentially associated with performance (e.g., Walkington, Clinton, & Shivraj, 2018). However, the environment also plays an essential role as it has been shown that teaching style and classroom management influence student performance (Greer, 1997). In the next sections task characteristics (i.e., linguistic, arithmetic factors), individual characteristics and environmental factors will be further described.

TASK CHARACTERISTICS

The task characteristics of a WP result from interrelated linguistic and arithmetic features (for summary see Daroczy et al., 2015). Importantly, both linguistic and arithmetic factors can be manipulated in such a way that the underlying mathematical problem, as well as the solver-built mental model, remains the same (i.e., isomorphic problems) or changes (i.e., non-isomorphic problems).

ISOMORPHIC PROBLEMS

Even simple changes to the text can either hinder problem comprehension or enhance performance (e.g., more concrete words; Walkington et al., 2018). Similarly, it is also possible to perform a mathematical manipulation, namely a change in numerical difficulty, and keep the text of the problem constant (Davis-Dorsey et al., 1991; Staub & Reusser, 1992; Stern & Lehrndorfer, 1992; Thevenot & Barrouillet, 2015; Vicente et al., 2007). The underlying mental model of an isomorphic problem remains unaffected by linguistic or mathematical manipulation.

Linguistic Factor. Nominalization is a factor that influences reading difficulty but does not change the underlying mathematical structure of a WP. Nominalization is a manipulated syntactic form that turns verbs into nouns, increasing the time needed to read a text (Gibson & Warren, 2004), that is found mostly in scientific works (MAK Halliday & Matthiessen, 2004). Nominalization also influences the solving process in WPs (Prediger et al., 2015). For example, if the text of a problem uses the nominalized form, “the selling”, instead of the verb form, “to sell”, text comprehension and student interest may decrease (Fang, 2006). In general, it seems that abstract (nominalized) forms of WPs may be more difficult for children to solve.

Arithmetic Factor. The presence of carry/borrow influences the difficulty of calculation but does not change the underlying mathematical structure of a WP. The text of the problem may remain unchanged as well. In carry problems, a 1 needs to be carried from the unit slot to the decade slot. Children and adults take longer and make more errors when computing the solution to a sum for which adding the units leads to a change in the number of tens (e.g., 14 + 9 = 23) (Deschuyteneer et al., 2005; Furst & Hitch, 2000) than when it does not (e.g., 11 + 12 = 23). We expect that WPs with carry/borrow will be more difficult to solve in this current study.

NON-ISOMORPHIC PROBLEMS

In non-isomorphic problems the manipulation of a linguistic or mathematical factor changes the underlying mathematical structure.

Linguistic Factors. WPs might contain specific keywords that hint at the solution (e.g., linguistic markers like “less” or “more”). In some cases, this is consistent with the required mathematical operation and in other cases it is not. This is referred to as lexical consistency or inconsistency. Studies have shown participants, especially children, display longer response times and higher reversal error rates when solving lexically inconsistent problems as compared to consistent problems (Hegarty et al., 1992; Hegarty et al., 1995; Pape, 2003; Van der Schoot et al., 2009; Verschaffel et al., 1992). Therefore, in the current study we expect participants to achieve lower solution ACC when solving lexically inconsistent problems.

Arithmetic Factor. Arithmetic operation is an arithmetic factor that changes the underlying mathematical structure of a WP and influences problem complexity (Koedinger & Nathan, 2004). As WPs requiring subtraction are harder to solve and reinforce a different strategy (De Corte & Verschaffel, 1987) we expect such problems in the current study will be of greater difficulty as compared to those requiring addition.

In sum, for the children in this study we expect both isomorphic and non-isomorphic linguistic and arithmetic factors to affect WP solution ACC.

INDIVIDUAL CHARACTERISTICS THAT INFLUENCE WORD PROBLEM SOLUTION ACCURACY

A wide range of studies have investigated individual characteristics that influence WP performance such as reading (e.g., Fuchs, Geary, et al., 2016), and mathematical skills (e.g., Thevenot & Barrouillet, 2015) working memory (e.g., Swanson & Fung, 2016), intelligence (e.g., Mori & Okamoto, 2017; Swanson & Beebe-Frankenberger, 2004) and other related factors like motivation (e.g., Gasco & Villarroel, 2014) and socio-economic status (e.g., Jordan et al., 2006).

LANGUAGE COMPETENCES

It has been shown that people with higher reading skill outperformed individuals with lower reading skill on WPs (Fuchs, Geary, et al., 2016; M. O. Martin & Mullis, 2013). It may be inferred that lower reading skill leads to lower solution ACC (Andersson, 2007; Chow & Ekholm, 2019; Jordan et al., 1995). Indeed, Fuchs et al. (2018) argue that solving a WP is a special form of text comprehension. Therefore, we expect that reading skill will have a significant influence on an individual's ability to solve WPs in this study.

MATHEMATICAL COMPETENCES

Thevenot and Barrouillet (2015) and Kail and Hall (1999) showed that besides reading, mathematical skill also has a significant impact on WP solution ACC. Some studies have found a direct effect of calculation skill on solution ACC when children solve WPs (e.g., Träff, 2013), while other studies did not find such an effect (e.g., Swanson & Fung, 2016). However, because there is converging evidence that even basic numerical processing skills are linked to higher mathematical achievement (De Smedt, Noël, Gilmore, & Ansari, 2013) we expect that mathematical competence will positively influence WP solution ACC in our study.

INTELLIGENCE

It has been shown that fluid intelligence predicts WP performance (Mori & Okamoto, 2017; Swanson & Beebe-Frankenberger, 2004). Thus, here we expect to replicate this effect.

WORKING MEMORY

The joint investigation of linguistic and numerical processes must consider joint moderator variables such as working memory capacity. It has been suggested that an individual's ability to solve WPs is generally related to working memory capacity (Passolunghi & Siegel, 2001) and a positive relationship between working memory capacity and solution ACC has been documented (Peng et al., 2016). Therefore, in this study, we expect working memory capacity to have a positive effect on WP solution ACC.

MOTIVATION

In the context of mathematics, motivation can be defined based on the expectancy value theory (e.g., Eccles, 2009). Within this framework, mathematical motivation is a three-dimensional construct with three subscales: individual appreciation, mathematical self-concept (i.e., belief about one's own mathematical ability) and interest. In general, motivation positively influences mathematical performance (Gaspard et al., 2016). This finding applies within the context of WPs as Gasco and Villarroel (2014) found a positive relationship between motivation and WP performance. For example, it has been shown that students who have a lower mathematical self-concept are unlikely to work on improving their problem-solving skills (Kloosterman & Stage, 1992). Thus, we expect in this study that motivation will influence performance on WPs.

SOCIO-ECONOMIC STATUS

Socio-economic status has been shown to be an important demographic factor in mathematical thinking development (Jordan et al., 2006), and a positive relationship between socio-economic status and WP solution ACC has been documented (e.g., Abedi & Lord, 2001). Therefore, we expect to see a positive relationship between WP performance and socio-economic status.

Task characteristics and individual characteristics are not the only factors that matter. Several studies indicate that the environment influences how students solve WPs differing in complexity and that the way WPs are taught in the classroom might influence strategy selection and performance (De Corte & Verschaffel, 1987; Lean et al., 1990).

ENVIRONMENTAL FACTORS

There are indications that in real life situations students are able to solve mathematical problems that they could not in the classroom (Terezinha Nunes Carraher, Carraher, & Schliemann, 1985). This contradicts the assumption that the failure of many students to engage in problem-solving behavior is attributable to low comprehension skills or cognitive deficits. Indeed failures might originate from sources other than cognitive deficits (Ulu, Tertemiz, & Peker, 2016), for example pedagogical factors. For instance, classroom organization, support offered to students and how solving WPs is explained has a great impact on learning (Titsworth et al., 2013). False assumptions about WPs (e.g, that every problem can be solved with a single operation; Cummins et al., 1988) are sometimes believed and instructed by teachers (Fuentes, 1998; Nortvedt, Gustafsson, & Lehre, 2016). One such assumption facilitated and even encouraged by teachers is the use of keyword strategies (e.g., Boote & Boote, 2018; Jitendra et al., 2015). Together, these findings indicate that teachers may have some influence over the fact that students pick out numbers and keywords instead of creating mental models (Hegarty et al., 1995). Of course, such strategies might work for simple problems, but not for more demanding ones. The keyword strategy is especially error-prone when applied to WPs containing lexically inconsistent (where the keyword does not refer to the correct operation) sentences (Hegarty et al., 1992; Hegarty et al., 1995).

Teaching method and quality of instruction are critical for proper learning and play an important role in academic performance (Thatcher et al., 2008). This applies not only to mathematical achievement in general, but also to WP success (Nortvedt et al., 2016). Finally, what happens in the classroom is a result of a complex social system (Lave, 1992), comprising teacher-student interactions, also referred to as classroom communication. The effects of classroom communication and quality of instruction on WP performance has not previously been directly investigated. Klieme et al. (2009) introduced a model to describe and measure teaching quality, which conceptualizes the interaction between students and teachers. Generally, it is unclear how teaching quality affects WP performance. In this study we expect that higher teaching quality to lead to greater solution ACC and better mathematical performance. Additionally, we should see a positive relationship between teaching quality scales with the more difficult problems. Teaching quality comprises three subscales: cognitive activation, supportive climate, and structured classroom management. The first subscale, cognitive activation integrates challenging tasks, the exploration of concepts, ideas and prior knowledge (Lipowsky et al., 2009). Therefore, in this study cognitive activation should be positively correlated with WP success because it encourages problem understanding, and not the use of the keyword strategy, prior to a merely stepwise, and simple rule-based, direct translation strategy. This means in classrooms that encourage cognitive activation we would hypothesize teachers use tasks that challenge students to create mental models. In such classrooms we expect that the better cognitive activation elicited decreases the inconsistency effect. Similarly, we also assume a positive effect on linguistic and arithmetic complexities. The second subscale of teaching quality, classroom management, focuses on classroom rules and procedures, coping with disruptions and smooth transitions (Fauth et al., 2014; Seidel & Shavelson, 2007). Student achievement is affected by classroom management, because classrooms that experience few disruptions, (i.e., classrooms with good classroom management) thus providing an environment where students are better able to concentrate on their tasks, are expected to be associated with WP solution ACC. The third subscale of teaching quality, supportive climate, covers “specific aspects of the teacher student relationship such as positive and constructive teacher feedback, a positive approach to student errors and misconceptions, and caring teacher behavior" (Fauth et al., 2014). Classrooms with a positive and supportive climate generally enhance student performance (Hugerat, 2016). Therefore, this subscale of teaching quality should have a positive impact on WP performance in this study because when students working with text problems are better supported and instructed by their teachers, they are likely to be better equipped to deal with difficulties.

One way of measuring teaching quality in primary schools are student ratings (Fauth et al., 2014; Sandilos et al., 2017). Considering student ratings is important because a student's view of their classroom reflects the actual characteristics of the classroom experience (Brock, Nishida, Chiong, Grimm, & Rimm-Kaufman, 2008; Howard, 2001). Moreover, results from Sandilos et al. (2017) indicated that student perceptions of teacher demands (challenge and control) are related to mathematical achievement.

Concluding, we can say that it is important to consider environmental factors, because they impact student performance. Additionally, task characteristics, individual characteristics and environmental factors have been tested in assorted studies either individually or together but very few studies have integrated all of these factors while distinguishing between arithmetically and linguistically simple and complex WPs.

RESEARCH QUESTIONS AND HYPOTHESES

The theoretical insights outlined above lead to the following research questions:

(1) How do individual characteristics influence performance?
- We expect that superior reading ability, mathematical skills, working memory and intelligence relate to better ACC when solving WPs.
- Additionally, higher motivation and socio-economic status should also increase WP solution ACC.

(2) How does environment act with linguistic and arithmetic difficulty to influence WP solution ACC?
- First, we expect that better teaching quality (classroom management, cognitive activation and supportive climate) lead to higher solution ACC and enhanced performance in mathematics.
- Additionally, we expect a positive relationship between teaching quality scales and arithmetically and linguistically more demanding items (cognitive activation in particular should increase ACC on lexically inconsistent items).

METHODS

PARTICIPANTS

The participants were 387 (Age M=10;7, SD=0.70) students from nine schools, 23 classes (gymnasiums) in the state of Baden-Württemberg Germany. All subjects were aged between 10 and 13 years (5th and 6th graders); 193 girls and 191 boys, three participants did not report their gender. Neurological or psychological disorders were exclusion criteria. Participation was on a voluntary basis.

STIMULI AND DESIGN

The study was a 2x2x2x2 within subject and orthogonal design with the main four factors: operation (addition vs. subtraction), carry/borrow (carry/borrow vs. non-carry/non-borrow), consistency (consistent vs. inconsistent form), nominalization (verbalized vs. nominalized form). Each of the 16 conditions consisted of 3 sentences, resulting in 48 WPs in total (see Study 3: Supplementary Material 1). The four factors were manipulated simultaneously and orthogonally, to examine not only the influence of single factors, which had not been examined thus far in WPs in isolation (linguistic and arithmetic complexity), but also their interaction. The problems were constructed in such a way that manipulation of one factor did not affect the other factors. To maintain solvability in terms of working memory our stimuli did not contain unnecessary information. All of our stimuli required mental calculation but not note taking. Each WP had the same form and included four sentences describing different stories. The first sentence was used as a reference point (e.g., *"A market woman is selling apples at the market"*). The second and third sentences contained the two numbers necessary for the arithmetic operation (e.g., *"The market woman arrived with 42 apples. She sold 17 apples").* The fourth and final sentence contained the question (e.g., *"How many apples did she have at the end?*"). Each task could be clearly solved in one calculation step; the text was not ambiguous and there was only one correct solution. All problems consisted of two-digit numbers in Arabic notation and overall problem size and parity (of results) was matched between non-carry and carry addition as well as between subtraction and addition problems. Ties, such as numbers with identical digits (e.g., 55), with full decades (e.g., 10, 20), all combinations of the operand where either tens or ones digits are the same (e.g., tens: 41 + 43; ones: 24 - 14), and mirror numbers (e.g., 24 - 42) were excluded to prevent any automatic mental retrieval. The second or the third sentence contained a cue word (in a verbalized or nominalized form) which evoked an operation (e.g., "sell"/"selling" for subtraction). Other words which might have provided hints towards the solution were specifically avoided (e.g., "in total"). The type of nominalization in the text was formulated in the so called perfect nominalized form. In this form words lose their verbal characteristics and behave like nouns (Hamm & van Lambalgen, 2002). An important aspect in preparing the stimuli was that the nominalized form was applicable to all selected keywords (verbs in our case) and that the changed form was as close to the "original" meaning as possible, therefore; all nominalized forms originated from the same lexical root as their corresponding verbs. All sentences were designed to be of comparable length in every task. All sentences were constructed using active verb forms as the active voice helps one to better understand WPs (Abedi et al., 2005). Filler words or information not relevant to understanding and solving the task were avoided to minimize the risk of adverse effects of prolonged reading and the extra burden of working memory on response time. There was a 15-minute time limit to solve all 48 of the WPs. The problems were divided into three 16 task blocks. Each block consisted of all conditions. Task order was pseudorandomised within each block, meaning that each condition could arise a maximum of two times. We avoided presenting the same order of WPs to participants sitting next to each other to discourage cheating.

MEASURES

Demographic Information. Demographic information was assessed using the questions of Hardy et al. (2011). Questionnaires also included queries about student *socio-economic background (SES)*. The SES questionnaire included 3 items. (1) How many books do you have at home; 0-10, 11-25, 26-100, 101-200, or over 200? (2) Which of the following items do you have at home; daily newspaper, dishwasher, learning program for computer, internet, lawnmower, and second automobile? Who lives together with you; mother, father, brother(s), sister(s), grandparents, or other? For the analysis, an aggregated scale was used: there was an overall score on a scale of 0 to 15 that increased with SES.

Cognitive Abilities.

Non-verbal intelligence was assessed with two scales (matrices and order) from the CFT 20-R (Weiß, 2006), an intelligence test unbiased by culture. Both scales contained 15 tasks and participants had limited time to solve them. For the analysis, the sum of all correct items was used (scale of 0-30).

Mathematical skill was assessed with two short speed calculation tests for arithmetic addition and subtraction (Huber et al., 2013). Both tests contained 28 math problems and no text. All problems consisted of two-digit numbers (number pairs were different from those used for WPs). Overall problem size and parity of the results was matched between non-carry and carry addition problems as well as for non-borrow and borrow subtraction problems. Carry/borrow and non-carry/non-borrow, as well as number pairs with even or odd results, were balanced.

The order was randomized. Participants had 90 seconds to solve each test. As an independent variable, we used the mean of the number of correctly solved items from the two tests (scale of 0 to 28).

Verbal memory span was measured with a simple memory span task consisting of 15 words. The experimenter read the words aloud and participants had to write down as many words as they could remember afterwards. For the analysis, the sum of all correctly and incorrectly remembered items was used in the following equation: 15 + (number of correct items – number of incorrect items). This resulted in a score on a scale of 0 to 30 that increased with verbal memory span.

Visual memory span was assessed with 30 various shapes presented sequentially using a projector. After the presentation ended, participants had to mark the projected shapes on an answer sheet they had seen before. The answer sheet contained 15 items from the presentation and 15 distractions. For the analysis, the sum of all correctly and incorrectly remembered items were used in the following equation: 15 + (number of correct items – number of incorrect items). This resulted in a score on a scale of 0 to 30 that increased with verbal memory span.

Reading abilities were measured using a speed of reading- and reading comprehension test designed for grades 6 through 12 (Schneider et al., 2007). After a short practice session, the subjects read a text for 4 minutes. The reading speed was calculated as the number of words read in the 4 minutes (scale of 0-1727). At certain points in the text word alternatives were offered and the task was to underline the word that fit the context. As an independent variable (reading comprehension) we used the number of correctly underlined words (2 * correct – incorrect). This resulted in a score on a scale of -23 to 64 that increased with reading comprehension.

Teaching quality. Teaching quality was assessed via student ratings based on the model of Klieme et al. (2009), where teaching quality comprises three basic dimensions: supportive climate, effective classroom management, and cognitive activation (Fauth et al., 2014; Sandilos et al., 2017). Student ratings were used to make a correlation between WP performance and teaching quality. Additionally, we measured the correlation between student perception of the classroom environment, solving WPs, and various other cognitive variables, such as motivation. Items for classroom management assessment were taken from Fauth et al. (2014). The questionnaire included 21 items on three sub scales: supportive climate (nine items), classroom management (five items), and cognitive activation (seven items). All items could be rated on a four-point scale ranging from strongly disagree to strongly agree (Table 1.). For all three scales we used the mean value of the responses (scale of 0 to 4) in the analysis.

Table 4 - 1 *Teaching Quality Scales.*

Scales	Items
Classroom management	
	In our science class …
	none of the students disturb the lesson
	students are quiet when the teacher speaks
	everybody listens, and students are quiet
	nobody interrupts with talking
	everybody follows the teacher
Cognitive activation	
	In our science class …
	we are working on tasks that I have to think about very thoroughly
	Our science teacher …
	asks me what I have understood and what I haven't
	asks questions that I have to think about very thoroughly
	gives us tasks that seem to be difficult at a first glance
	asks what we know about a new topic
	gives us tasks I like to think about
	wants me to be able to explain my answers
Supportive climate	
	Our science teacher …
	is nice to me even when I make a mistake
	cares about me
	encourages me when I find a task difficult
	tells me how to do better when I make a mistake
	likes me
	tells me what I'm already good at and what I still have to learn
	is friendly to me compliments me when I did something good
	believes that I can solve difficult tasks

Motivation. The motivation questionnaire included 13 items on three sub scales: self-concept (5 items with 3 inverse items; e.g., "I am good in mathematics" or "I am not that good in mathematics"), interest (3 items with 1 inverse item; e.g., "dealing with mathematics is one of my favourite things to do”, or “I am not interested in mathematics”), individual appreciation (5 items with 2 inverse items; e.g., “for me it is important to know a lot in mathematics”, or “mathematics is a real burden to me”) adapted from Gaspard et al. (2016). All items were rated on a four-point scale ranging from strongly disagree to strongly agree (See Table 2). For all three scales we used the mean value of the response (scale of 0 to 4) in the analysis.

Table 4 - 2 *Motivation Scales.*

Scale	Item
Self-concept	
	I am good at math.
	I simply have no talent for mathematics.
	I do not like mathematics.
	Mathematics is easy for me.
	I always have problems with math problems.
Interest	
	Dealing with mathematics is one of my favorite activities.
	I am not interested in mathematics.
	I enjoy working on books or brain teasers related to math.
Individual Importance	
	I do not like to do mathematics.
	It is important to me to know a lot in mathematics
	It is important to me to be good at math.
	Mathematics is a real burden for me.
	Mathematics is very useful to me

PROCEDURE

Children were recruited from schools in the state of Baden-Württemberg. After informing the head of the school about the study and obtaining provisional participation agreement, we recruited children willing to participate. Parents or legal guardians, teachers and children were informed about the study. If a child and their legal guardians agreed to participate in the study the measurement took place during two class hours (ca. 90 min) in the form of group testing at the schools. Informed consent and some basic information were obtained from a legal guardian as well as the student prior to participation. All tests were carried out using paper and pencil. Data were collected during classroom-wide assessments by trained staff using standardized instructions.

RESULTS

Two separate analyses were conducted. The first was an individual level analysis to investigate the association between individual abilities and WP solution ACC without environmental factors. The second, was an effect-based analysis to investigate the association between environmental factors, individual abilities and task characteristics. Raw means and standard deviations are shown in Table 3. ACC results were arcsine-square-root-transformed prior to analysis to approximate a normal distribution (see Winer, Brown, & Michels). Individual variables were all z-transformed for better interpretability and comparability of different scales.

Table 4 - 3 *Descriptive Data (Raw Means, SD, Median, Skewness).*

		Total Sample N=359				Restricted Sample N=279			
		Mean	SD	Median	Skewness	Mean	SD	Median	Skewness
Demographic Data									
	Age	10.86	0.71	11	0.17	10.92	0.71	11	0.17
	Socio-economic Status	11.5	2.11	12	-0.53	11.63	2.12	12	-0.6
Cognitive Measures									
	Intelligence	22.28	3.5	23	-0.5	22.6	3.42	23	-0.53
	Reading Speed	582.91	242.12	556	1.82	599.51	227.82	570	1.72
	Reading Comprehension	8.19	4.22	8	0.39	8.58	4.14	8	0.44
	Visual Span	25.15	3.06	25	-2.05	25.24	3.14	26	-2.41
	Verbal Span	23.12	2.85	23	-3.02	23.12	3.03	23	-3.25
	Mathematic Performance	22.58	3.96	22.75	-0.59	23.51	3.44	23.75	-1.02
Classroom Measures L1									
	Cognitive Activation L1	2.85	0.53	2.8	-0.5	2.82	0.55	2.8	-0.5
	Classroom Management L1	2.24	0.66	2.17	-0.11	2.26	0.64	2.2	-0.17
	Supportive Climate L1	3.28	0.56	3.44	-1.25	3.28	0.57	3.44	-1.37
Classroom Measures L2									
	Cognitive Activation L2	2.85	0.19	2.92	-0.67	2.84	0.2	2.89	-0.54
	Classroom Management L2	2.24	0.33	2.35	0.06	2.26	0.34	2.35	0
	Supportive Climate L2	3.28	0.23	3.32	-0.67	3.28	0.23	3.3	-0.6
Motivation									
	Motivation: Interest	2.55	0.8	2.67	-0.18	2.6	0.79	2.67	-0.24
	Motivation: Individual Importance	3.2	0.63	3.2	-0.79	3.24	0.64	3.4	-0.95
	Motivation: Self-Concept	3.11	0.67	3.2	-0.84	3.18	0.67	3.2	-1.08
Performance Measures									
	Total Accuracy (percentage)	0.77	0.19	0.77	-0.47	0.84	0.13	0.85	-0.26
	Accuracy B1+B2 (percentage)	0.82	0.16	0.88	-1.31	0.89	0.09	0.91	-0.98
Linguistics Effects									
	NomEffB2	-0.34	1.72	0	0.09	-0.37	1.65	0	-0.02

	ConEffB2	-0.82	1.96	-1	-0.32	-0.81	1.7	-1	-0.33
Mathematical Effects									
	CarEffB2	-0.91	2.1	-1	-0.99	-0.7	1.84	0	-0.89
	OpEFFB2	-0.7	2.02	-1	-0.17	-0.63	1.84	-1	-0.04

First, in both types of analysis (individual or effect based) we calculated the intraclass correlation coefficients (ICCs; see Table 4) to define the proportion of total variance that could be attributed to between class differences. ICC(1) is the index of the average agreement between pairs of students within the same class, whereas ICC(2) refers to the reliability of the group average (Marsh et al., 2012). In the case of student ratings, ICC(1) can be seen as the degree to which student ratings of classroom features are affected by being in different classes (Lüdtke, Robitzsch, Trautwein, & Kunter, 2009). In multilevel studies, ICC(1)s for climate variables are often less than .10 and rarely greater than .30 (e.g., Marsh, Martin, & Cheng, 2008).

Table 4 - 4 *ICC(1) & ICC2 Values.*

		ICC(1)	ICC2	ICC(1)	ICC2
Demographic Data					
	Socio-economic Status	0.03	0.34	0.05	0.39
Cognitive Measures					
	Intelligence	0.07	0.53	0.10	0.57
	Reading Speed	0.04	0.37	0.01	0.13
	Reading Comprehension	0.10	0.62	0.06	0.44
	Visual Span	0.06	0.48	0.06	0.44
	Verbal Span	0.11	0.64	0.08	0.50
	Mathematics Performance	0.08	0.55	0.11	0.58
Classroom Measures L1					
	Cognitive Activation L1	0.07	0.54	0.08	0.52
	Classroom Management L1	0.21	0.80	0.24	0.78
	Supportive Climate L1	0.11	0.64	0.10	0.58
Motivation					
	Motivation: Interest	0.04	0.38	0.10	0.56
	Motivation: Individual Importance	0.05	0.43	0.09	0.53
	Motivation: Self-Concept	0.06	0.48	0.11	0.59
Performance Measures					
	Total Accuracy	0.15	0.72	0.03	0.26
Linguistics Effects					
	Nominalization Effect	-	-	0.05	0.38

	Consistency Effect	-	-	0.00	-0.01
Mathematical Effects					
	Carry/borrow effect	-	-	0.02	0.23
	Operation Effect	-	-	-0.03	-0.62

For the *individual level analysis*, we used all three WP blocks (i.e., all 48 WPs) and we included all subjects who completed the questionnaires and cognitive tasks in full. Out of 387 participants, 359 (age M=10.87, SD=0.71) students were included in this analysis: 179 girls (age M=10.87, SD=0.69) and 180 boys (M=10.84, SD=0.73). At the individual level ICC(1) indicated that 15 % of the variance in total ACC could be explained by classroom differences. In *Model 0*, intelligence, reading speed, reading comprehension, mathematic performance, verbal/visual span, motivational scales and SES were entered as predictors. The clustering at the classroom level was handled by the TYPE = COMPLEX option in Mplus (see Muthen & Satorra, 1995).

For the *effect-based analysis* we used the first two WP blocks (B1+B2) containing 32 problems. Participants who had missed more than 3 continuous items and those for whom more than 20 % of the data was missing were excluded from the analysis blocks. With these restrictions 273 (age M=11;7, SD=0.74) students from 23 classes were included in the effect-based analysis: 141 girls (age M=11;7, SD=0.68) and 132 boys (age M=11;7, SD=0.84). Difference scores in ACC were generated between the following: inconsistent tasks and consistent tasks (*consistency effect*); nominalized tasks and verbalized tasks (*nominalization effect*); subtraction tasks and addition tasks (*operation effect*); carry/borrow tasks and non-carry/non-borrow tasks (*carry/borrow effect*). These effects can be interpreted similarly, taking for example nominalization: a positive value for the nominalization effect means that participants experience more failure in nominalized tasks and higher values indicate more errors due to the difficulty of the nominalization. The difference between effect scores reflects differences in performance on the different WPs. In the effect based analysis ICC(1) values indicated that 2 % and 5% of the total variance may have been due to classroom differences for the *nominalization effect* and the *carry/borrow effect*, respectively. ICC(1) values for the *operation effect* and *consistency effect* were negative or zero. This means that there was no level 2 variability in the outcomes when accounting for classroom membership; therefore, we did not conduct the multilevel analysis. In the nominalization and carry/borrow effect models, students (level 1; L1) were nested in classrooms (level 2; L2). Clustering at the classroom level was handled by the TYPE = TWOLEVEL option in Mplus. This way we examined four models each for nominalization and the carry/borrow effect. The basic dimensions of teaching quality were included separately (*Models 1-3*) as well as together (*Model 4*).

INDIVIDUAL PREDICTORS

In *Model 0*, intelligence, reading speed, reading comprehension, and mathematic performance were significant positive predictors of total ACC. Visual span was a negative predictor while motivational scales, SES and verbal span were not significant predictors of total ACC. *Model 0* explains 47% of the variance in total ACC. Mathematical performance had the highest impact on total ACC followed by reading speed, reading comprehension and then intelligence. For the result outputs of *Model 0* see Table 5.

Table 4 - 5 *Individual Level of Analysis.*

Predictor	Model 0
Intelligence	0.08(0.02) *
Reading Speed	0.11 (0.03) *
Reading Comprehension	0.42 (0.14) *
Visual Span	-0.06 (0.02)*
Verbal Span	-0.08 (0.04)
Mathematics	0.39 (0.05)*
Interest	0.04 (0.04)
Individual Importance	-0.02 (0.05)
Self Concept	-0.01 (0.05)
Socio-economical Status	0.04 (0.02)
R^2	0.47

EFFECT BASED ANALYSIS

Nominalization effect. Intelligence, reading comprehension and self-concept were significant predictors at L1 in each of the models, and accounted for 6% of the variance. The effects of nominalization were smaller with higher intelligence and reading comprehension and mathematics self-concept magnified the effect of nominalization. In Model 1 supportive climate accounted for 63% of the L2 variance. In *Model 2* cognitive activation accounted for 96% of the L2 variance. In *Model 3* intelligence, reading comprehension and self-concept were significant predictors at L1, and accounted for 6% of the variance. In *Model 4* cognitive activation and supportive climate ($p < 0.1$) together accounted for 97 % of the L2 variance (see Table 6).

Table 4 - 6 *Nominalization Effect.*

	Model 1	Model 2	Model 3	Model 4
Individual Level				
Intelligence	-0.09 (0.03)*	-0.09 (0.03)*	-0.09 (0.03)*	-0.09 (0.03)*
Reading Speed	0.04 (0.04)	0.04 (0.04)	0.04 (0.04)	0.04 (0.04)
Reading Comprehension	-0.09 (0.04)*	-0.09 (0.04)*	-0.09 (0.04)*	-0.09 (0.04)*
Visual Span	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)
Verbal Span	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)
Interest	-0.07 (0.06)	-0.07 (0.06)	-0.07 (0.06)	-0.07 (0.06)
Individual Importance	-0.02 (0.08)	-0.02 (0.08)	-0.02 (0.08)	-0.02 (0.08)
Selfconcept	0.14 (0.06)*	0.14 (0.06)*	0.14 (0.06)*	0.14 (0.06)*
Socio-economic Status	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)
Mathematic	0.04 (0.04)	0.04 (0.04)	0.04 (0.04)	0.04 (0.04)
Classroom Level				
Supportive Climate	-0.07 (0.04)*			-0.05 (0.03)
Cognitive Activation		-0.11 (0.02)*		-0.08 (0.03)*
Classroom Management			0.04 (0.03)	0.04 (0.03)
R^2 within	0.07	0.07	0.07	0.07
R^2 between	0.92	0.93	0.95	0.95

* two-tailed significant

Supportive Climate p<0.1

Carry/borrow effect. Intelligence was a significant L1 predictor. There were no predictors found at L2 (see Table 7).

Table 4 - 7 *Carry/borrow Effect.*

	Model 1	Model 2	Model 3	Model 4
Individual Level				
Intelligence	0.08 (0.04)*	0.08 (0.04)*	0.08 (0.04)*	0.08 (0.04)*
Reading Speed	0.01 (0.03)	0.01 (0.03)	0.01 (0.03)	0.01 (0.03)
Reading Comprehension	-0.02 (0.03)	-0.02 (0.03)	-0.02 (0.03)	-0.02 (0.03)
Visual Span	-0.02 (0.05)	-0.02 (0.05)	-0.02 (0.05)	-0.02 (0.05)
Verbal Span	-0.04 (0.04)	-0.04 (0.04)	-0.04 (0.04)	-0.04 (0.04)
Interest	-0.02 (0.06)	-0.02 (0.06)	-0.02 (0.06)	-0.02 (0.06)
Individual Importance	0.09 (0.08)	0.09 (0.08)	0.09 (0.08)	0.09 (0.08)
Selfconcept	-0.05 (0.04)	-0.05 (0.04)	-0.05 (0.04)	-0.05 (0.04)
Socio-economic Status	-0.04 (0.04)	-0.04 (0.04)	-0.04 (0.04)	-0.04 (0.04)

Mathematic	0.05 (0.05)	0.05 (0.05)	0.05 (0.05)	0.05 (0.05)
Classroom Level				
Supportive Climate	0.00 (0.04)			-0.04 (0.05)
Cognitive Activation		-0.01 (0.04)		0.02 (0.05)
Classroom Management			0.07 (0.04)	0.08 (0.04)
R^2 within	0.05	0.05	0.05	0.05
R^2 between	0.02	0.01	0.17	0.38

In sum, 15 % of the differences in total ACC, as well as 5% of the nominalization and 2% of the carry/borrow effects, can be explained by classroom differences. The *individual level analysis*, which explained 49% of the differences in total ACC, showed that intelligence, reading speed, reading comprehension and mathematical skill account for greater solution ACC, whereas visual span accounted for lower solution ACC. *The effect-based analysis* revealed that for the carry/borrow effect there was no significant L2 predictor, and only intelligence turned out to be a significant L1 predictor. The nominalization effect had less of an impact when students rated supportive climate and cognitive activation highly. Finally, L1 predictors such as high intelligence and reading comprehension also accounted for better results in the face of nominalization.

DISCUSSION

This article presents assessments of WP performance in relation to individual abilities and environmental factors. First, we replicated previous findings that individual abilities play a role in problem-solving skill. However, not all predictors were found to contribute significantly to performance. Substantively, we were able to confirm cognitive activation to be predictive of performance when WPs are presented in nominalized forms. Theoretically, our results provide evidence for the importance of teaching quality on WP performance.

INDIVIDUAL CHARACTERISTICS THAT INFLUENCE WORD PROBLEM SOLUTION ACCURACY

We expected better reading ability, mathematical skills, working memory and higher intelligence and motivation to positively influence student ability to accurately solve WPs. The *individual level analysis* showed that higher intelligence, reading speed, reading comprehension and mathematical skill scores account for greater solution ACC, additionally visual span accounts for lower solution ACC.

Reading Skill. As expected, reading skill accounted for better solution ACC. This is in line with several studies (e.g., Fuchs, Geary, et al., 2016; Mullis, Martin, & Foy, 2011) which showed that reading skill directly influenced the solving of WPs and argued for a strong role of language comprehension skills. This is opposite to the finding of Lee, Ng, and Bull (2017) who found language skill did not discriminate between WP difficulty levels (Lee et al., 2017). Although, Lee et al. (2017) pointed out that it is possible that the WPs presented failed to capture a broad range of language demand. However, our results from the effect-based analysis show that reading comprehension also accounted for a reduction in the nominalization effect. This indicates that good reading competencies aid problem-solving, especially if the linguistic part of the WP is difficult.

Mathematical skill. As expected, mathematical skill contributed to higher WP solution ACC. This is in line with some previous research indicating mathematical skill is a necessary foundation for solving WPs (e.g., Träff, 2013) but contradicts the results of other studies that found no effect on problem-solving ACC (e.g., Fuchs, Gilbert, et al., 2016; Swanson & Fung, 2016). The reason we obtained the results that we did might be due to how we measured mathematical ability. As we used speed calculation tests, our design allowed for those who could calculate faster to solve more problems, therefore; participants who solved more problems within the given time had higher total ACC values. Additionally, in the presence of the nominalization effect, mathematical skill was not a significant predictor, but there was a correlation between mathematical skill, language skill and motivation, which indicates a relationship between these factors. These results are supported by Swanson (2004), who found that calculation skills mediated WP solution achievement through reading skill and intelligence.

Intelligence. As expected, intelligence accounted for higher WP solution ACC and was a significant predictor at L1 when the teaching quality scales were added to *Models 2-4*. This is in line with previous research suggesting a direct relationship between intelligence and WP solving (Mori & Okamoto, 2017; Swanson & Beebe-Frankenberger, 2004). Greater intelligence accounted for higher ACC in the face of the nominalization effect, which indicates that children of higher intelligence made fewer errors when solving problems with nominalization.

Finally, in the case of the carry/borrow effect only intelligence turned out to be a significant negative L1 predictor, likely meaning that children of higher intelligence make more calculation errors. This is unexpected as we anticipated a smaller carry/borrow effect for children of higher intelligence, as such children usually perform better on more complex tasks compared to children with normal intelligence (Xin & Zhang, 2009). A reason why we obtained such a result could be a speed-ACC trade-off; that is, highly intelligent children solved the problems quickly but at the cost of making more mistakes. There was a weak positive correlation between the number of fully solved items and the number of errors ($r(359) = .16$, p

= .003) in the total sample. Indeed, in the carry/borrow effect analysis we included only those participants who had answered the first 32 WPs, and not only the first 16 WPs. This criterion reduced our sample size by 3/4 (i.e., 100 children were excluded), so theoretically the fastest children were included in this analysis. However, there was no significant correlation between intelligence and the carry/borrow effect; therefore, these results should be interpreted with care.

Memory. Verbal span did not turn out to be a significant predictor of solution ACC. It has been suggested that WP skills are in general related to working memory (Passolunghi & Siegel, 2001; Peng et al., 2016). Even though most studies have found a strong relationship between memory components and solving WPs (see for a reviewPeng et al., 2016) the existence of such a relationship is uncertain. Some studies did not find a direct effect of working memory (Kail & Hall, 1999; Swanson et al., 1993) especially when reading and calculation skills were taken into account. One potential reason why we did not find verbal span to be a significant predictor of solution ACC in this study is we measured verbal and visuospatial short-term memory. Because there is evidence that working memory is likely more strongly involved in WP processing than short term memory components (Passolunghi & Costa, 2019).

Visual span was a significant predictor of ability to solve WPs, but higher values in visual span resulted in lower total ACC. On the other hand we found positive correlations between visual/verbal memory and motivational scales, language competences, intelligence and mathematics. First, this is in line with other studies that found memory to be related to other individual factors. For instance, general language comprehension has been shown to be mediated by working memory for the specific language used in WPs (Andersson, 2007; Fuchs, Geary, et al., 2016). Additionally, intelligence is also highly correlated with working memory (Ackerman, Beier, & Boyle, 2005). However, our results hint towards collinearities, where visual span becomes a suppressor variable. This is supported by the fact that visual span had a zero correlation with ACC, but correlated positively with predictor variables such as intelligence, mathematics, and reading comprehension, and correlated negatively with reading speed.

Motivation. The three scales of motivation did not affect total ACC. Nevertheless, there were positive correlations between motivational scales and reading comprehension, as well as between mathematical skill and total performance. Unexpectedly the sub-scale “mathematical self-concept” turned out to be a significant negative predictor for the nominalization effect. This means that children with higher self-concept achieved lower ACC when solving WPs with nominalization. This result indicates that higher self-concept results in a greater number of errors when solving WPs containing sentences with nominalization as compared with verbalization. This was somewhat unexpected, because higher self-concept should predict better performance (OECD, 2013). However, the zero-order correlation between self-concept and the nominalization effect was weak in the limited sample included in the effect-based analysis, indicating that our data does not provide enough evidence for the existence of an association between self-concept and the nominalization effect.

Socio-economic status. Contrary to the findings of the OECD (2013), which showed that socio-economically advantaged students have higher scores in mathematics, we found no effect of SES in the first analysis. One explanation is that the socio-economic question did not differentiate enough between students, as the entire group was in a gymnasium school made up of students with comparable socio-economic backgrounds. Another explanation is that according to the OECD (2013), the impact of SES on problem-solving performance was weaker than on performance in mathematics, reading or science. However, in the effect-based analysis, SES was a significant L1 predictor for the nominalization effect, meaning that higher SES led to better ACC on the linguistically more demanding WPs. SES has been related to language development in childhood (Pace, Luo, Hirsh-Pasek, & Golinkoff, 2017), and associated with better language abilities (Calvo & Bialystok, 2014). Therefore, one possible explanation for our results is that children from a higher socio-economic background are more familiar with complicated language containing nominalization.

ENVIRONMENTAL FACTORS

The hypothesis that teaching quality affects solution ACC on arithmetically and linguistically complex items (i.e., when the task characteristics differ) was partially confirmed. Namely, we showed that higher quality teaching (cognitive activation, supportive climate) leads to higher ACC in the presence of linguistically demanding items like nominalization. Nortvedt et al. (2016) found instructional quality significantly predicted mathematical achievement and reading comprehension in a number of countries

Model 1 and Model 2 showed that supportive climate and cognitive activation reduced the effect of nominalization. In Model 4 where all teaching quality scales were included, cognitive activation alone reduced the effect of nominalization. It has been shown that denser nominalization generally results in lower solution ACC (Prediger et al., 2015), but our results indicate that students from classrooms with better cognitive activation can solve linguistically demanding tasks. This is important because high quality instruction might weaken the relationship between reading comprehension and mathematical achievement (Adelson, Dickinson, & Cunningham, 2015).

The effect-based analysis revealed that for the carry/borrow effect there was no significant L2 predictor. Nevertheless, the ICC1 revealed that only 2% of the carry/borrow effect could be explained by classroom differences. This indicates that classes did not differ regarding the carry/borrow effect. In addition, contrary to expectations, we found no differences between classrooms in teaching quality either for lexical consistency or for operation. This means that there was no between classroom variation for lexical consistency and operation. Although, we did not specifically assess how the teachers taught WPs, and to what extent their teaching material contained inconsistent examples, this result indicates that solving WPs with carry/borrow, addition/subtraction, or with consistent/inconsistent forms might be taught similarly across different classrooms.

Lastly we expected, based on the results of Fauth et al. (2014), better classroom management to lead to better solution ACC on more difficult WPs. However, classroom management was not a predictor in any model. However, classroom management would not be the first indicator to predict effect-based results that are a very specific indicator of student ability to cope with difficult text in WPs. This means that this subscale of teaching quality is not that important when compared to the other subscales. This indicates that whether students are able to deal with complex WPs probably depends more on the way mathematics is instructed in their classroom and if proper support is offered during the learning process rather than on the general level of classroom management. Nevertheless, our results show that aggregated classroom management value correlated positively with total performance.

In sum, we found no evidence that the teaching quality plays a role in the case of lexical consistency, the carry/borrow effect or operation. For these factors, it seems that there is little or no differences between classrooms. However, better classroom cognitive activation reduces the nominalization effect.

LIMITATION

At the individual level, 49% of the variance could be explained, meaning that other, additional linguistic and cognitive factors play a role. For instance, emotional factors such as anxiety have been shown to play a considerable role in WP performance (Passolunghi, Cargnelutti, & Pellizzoni, 2019), or a connection with individual interests might increase solution ACC. For example, Mattarella-Micke and Beilock (2010) found increased errors on multiplication problems containing irrelevant numerical information when the irrelevant numbers were associated, rather than unassociated, with the protagonist of the WP (in this case, association means that if the protagonist picks up an object, it becomes associated with the protagonist whereas if they put down an object, it becomes unassociated with the protagonist). This example also shows that the wording surrounding numerals is of crucial importance and an area where additional research is needed. In this paper we investigated factors which were either linguistically/numerically isomorphic or non-isomorphic (i.e., manipulating the factor does or does not change the underlying mathematical structure); however, we did not consider a factor, like a linguistic change, which does not change the underlying structure but manipulates the calculation process. For example, Bassok et al. (1998) have shown that in WPs additions are easier when people have to add up functionally connected words (e.g., bird-cages) than categorically connected words (e.g., tulips, daisies) Additionally, (Bagnoud et al., 2018) found brain activity differed when solving WPs describing discrete quantities (e.g. sand), and WPs describing continuous quantities (e.g. apples, meter, rope), which indicates a deeper connection between linguistic and arithmetic factors.

THEORETICAL IMPLICATIONS

In this study, we demonstrated that both environmental factors and individual abilities affect performance on WPs. Our results show that performance does not only depend on reading comprehension, mathematical skills, and intelligence, but also on how supportive the environment is. Our finding that cognitive activation and supportive climate reduced the nominalization effect means that, for instance, positive and constructive teacher feedback, a positive approach to errors and misconceptions, the integration of challenging tasks and the exploration of concepts and ideas help to achieve better results with linguistically complex WPs.

The educational implications of our findings are relevant to the research and have a bearing on practical issues. This study provides evidence that teaching quality influences WP performance. This is important because Nortvedt et al. (2016) showed that only 12 % of teachers offered high quality instruction in both reading and mathematics, although mathematical texts are claimed to require special literacy skills (e.g., Burton and Morgan (2000) or Österholm and Bergqvist (2012)). Recent findings argue the importance of drawing student attention not only to mathematical but also linguistic aspects of WPs (Fuchs et al., 2018; Kyttälä & Björn, 2014). This suggests closer attention should be paid to building an association between WP and numerical cognition research.

SUMMARY

This article presents assessments of WP solving in relation to individual abilities and teaching quality. The first goal of this study was to investigate individual characteristics and how these affect the solution process. The *individual level analysis* showed that higher intelligence, greater reading speed and reading comprehension and better mathematical skill account for greater solution ACC. Higher intelligence and greater reading comprehension accounted for better results overall, and better results in the face of nominalized sentences, too. These results are in line with previous findings. Secondly, we also investigated how teaching quality (classroom management, cognitive activation and supportive climate) affects WP solution ACC while distinguishing between arithmetically and linguistically simple and complex WPs. *The effect-based analysis* revealed that a more supportive climate and higher cognitive activation reduces the nominalization effect. Theoretically, our results provide evidence that besides the importance of individual factors, environmental factors affect performance on linguistically demanding WPs.

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments.

SUPPLEMENTARY MATERIAL: MAIN EFFECTS AND INTERACTIONS

MAIN EFFECTS

Accuracies were evaluated by a 2 × 2 ANOVA, and tested pair-wise. 273 students were included into the analysis. In the analysis, the partial eta-squared (η^2_p) was calculated as a measure of effect size (Pierce et al., 2004). According to Pierce et al. (2004), values of 0.02, 0.13, and 0.26 represent small, medium, and large effect sizes respectively. For the descriptive data see Table 1.

Table 4S - 1 *Descriptive Data for Behavioral Measures.*

	Addition				Subtraction			
	Non-Carry		Carry		Non-Carry		Carry	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Consistent Form								
Verbalised Form	0.88	0.24	0.83	0.29	0.89	0.24	0.82	0.28
Nominalised Form	0.84	0.27	0.82	0.29	0.84	0.27	0.85	0.26
Inconsistent Form								
Verbalised Form	0.85	0.26	0.83	0.29	0.81	0.28	0.74	0.34
Nominalised Form	0.86	0.26	0.82	0.28	0.81	0.29	0.64	0.34

Nominalization resulted in a main effect $F(1,272)=10.75$ $p=.001$ $\eta_p^2=.04$, indicating, that sentences with nominalized form result in lower ACC compared to verbalized form. Sentences with verbalized form (90.02 %) had higher ACC than sentences with nominalized form (87.94 %).

Lexical Consistency resulted with the highest effect size in a main effect $F(1,272)=69.90$ $p<.001$ $\eta_p^2=.20$, indicating, that sentences with lexically consistent form (91.60 %) had higher ACC than sentences with lexically inconsistent form (86.35 %).

Carry/Borrow resulted with a medium effect size in a main effect $F(1,272)=34.81$ $p<.001$ $\eta_p^2=.11$, indicating, that sentences with non-carry/non-borrow (91.00 %) had higher ACC than sentences with carry/borrow (86.95 %).

Operation resulted with a medium effect size in a main effect $F(1,272)=46.14$ $p<.001$ $\eta_p^2=.15$, indicating, that sentences with addition (91.00 %) had higher ACC than sentences with subtraction (86.95 %).

INTERACTIONS

Accuracies were evaluated by a 2 × 2 ANOVA, and tested pair-wise. 273 students were included into the analysis.

Nominalization * Lexical Consistency. We observed main effects for both factors: Nominalization $F(1,272)=11.10$ $p=.001$ $\eta_p^2=.04$, Consistency $F(1,272)=69.26$ $p<.001$ $\eta_p^2=.20$. No interaction between the two factors were found $F(1,272)=3.00$ $p=.085$ $\eta_p^2=.01$ (see Figure 2).

Nominalization * Operation. We observed main effects for both factors. Nominalization $F(1,272)=10.23$ $p=.002$ $\eta_p^2=.04$, Operation $F(1,272)=45.29$ $p<.001$ $\eta_p^2=.14$. Additionally, a two-way interaction between the factors have been found $F(1,272)=6.00$ $p=.015$ $\eta_p^2=.02$. The post-hoc test revealed that there was no significant interaction between verbalized form and nominalized form in the case of addition $F(1,272)=0.29$ $p=.593$ $\eta_p^2=.00$, but in the case of subtraction $F(1,272)=15.66$ $p<.001$ $\eta_p^2=.05$. This indicates that the extra linguistics difficulty arises in the combination with the mathematically more complex case (see Figure 2).

Nominalization * Carry/Borrow. We observed main effects for both factors. Nominalization $F(1,272)=12.55$ $p<.001$ $\eta_p^2=.04$, Carry/Borrow $F(1,272)=36.46$ $p<.001$ $\eta_p^2=.12$. No interaction between the two factors were found $F(1,272)=.92$ $p=.339$ $\eta_p^2=.03$ (see Figure 3).

Operation * Carry/Borrow. We observed main effects for both factors. Operation $F(1,272)=49.23$ $p<.001$ $\eta_p^2=.15$. Carry/Borrow $F(1,272)=34.86$ $p<.001$ $\eta_p^2=.11$. Furthermore, there was a significant interaction between the two factors $F(1,272)=9.85$ $p=.002$ $\eta_p^2=.04$.

The post-hoc test revealed that there is a significant difference between carry and non-carry in both case of addition $F(1,272)=4.83$ $p=.029$ $\eta_p^2=.02$ and subtraction $F(1,272)=45.73$ $p<.001$ $\eta_p^2=.14$. The results hint at an over additive interaction between the factors, namely that the carry effect is especially pronounced in the case of subtraction (see Figure 1).

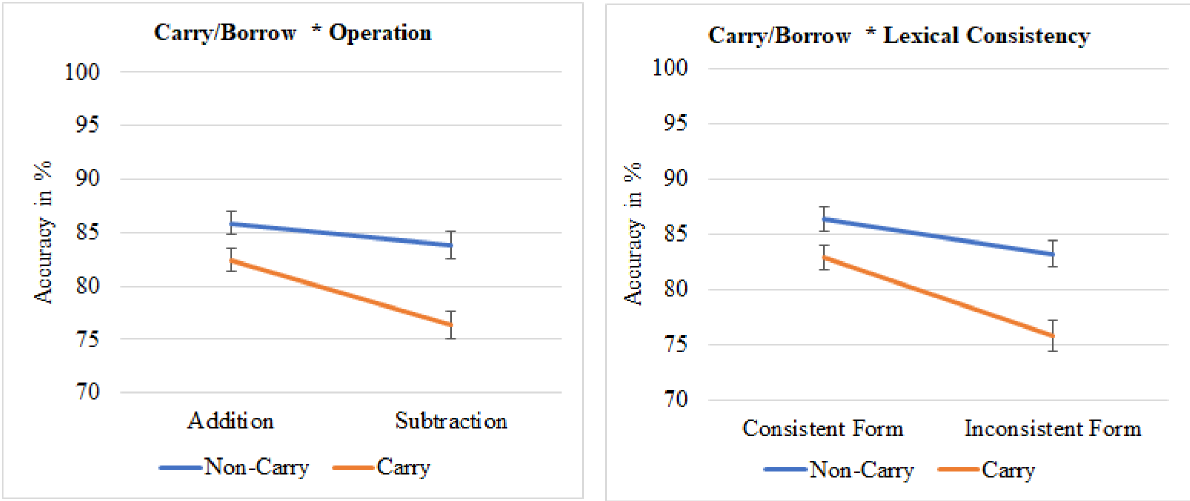

Figure 4S - 1 Interaction of Carry/borrow * Operation and Carry/borrow * Lexical consistency.

Carry/Borrow * Consistency. We observed main effects for both factors. Carry/Borrow F(1,272)=32.47 p<.001 η_p^2=.11, Consistency F(1,272)=68.80 p<.001 η_p^2=.20. Furthermore, there was a significant interaction between the two factors F(1,272)=4.96 p=.027 η_p^2=.02. The post-hoc test revealed that the consistency effect could be observed in both non-carry/borrow F(1,272)=20.04 p<.001 η_p^2=.07 and carry/borrow sentences F(1,272)=57.43 p<.001 η_p^2=.17. The results hint at an over additive interaction between the factors, namely that the carry effect is especially pronounced in the case of lexically inconsistent cases (see Figure 2).

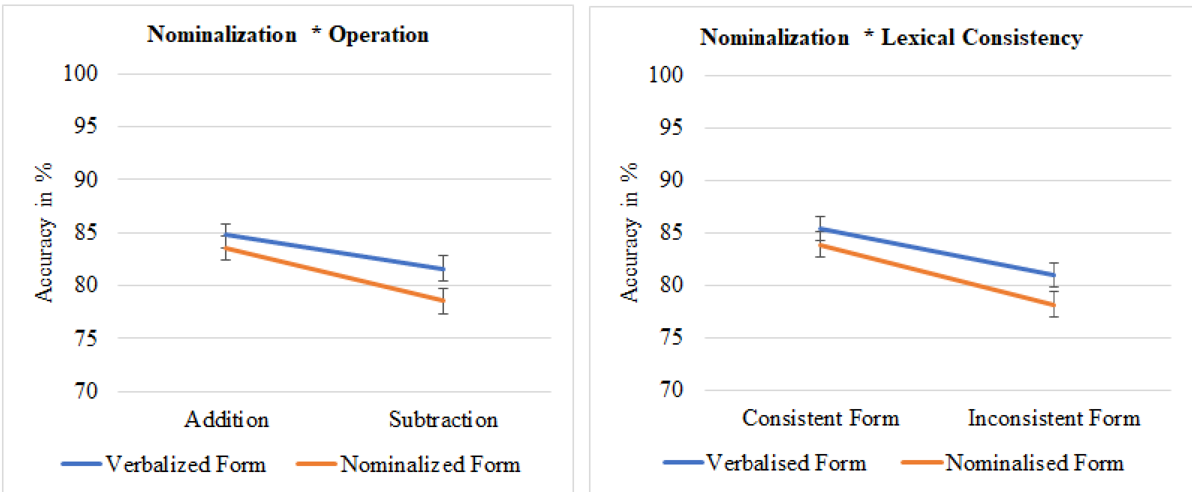

Figure 4S - 2 Interaction of Nominalization * Operation and Nominalization * Lexical consistency.

Operation * Consistency. We observed main effects for both factors. Operation F(1,272)=38.11 p<.001 η_p^2=.12. Consistency F(1,272)=57.54 p<.001 η_p^2=.18. Furthermore, there was a significant interaction between the two factors $F(1,272)=35.91$ $p<.001$ $\eta_p^2=.12$. The post-hoc test revealed that the consistency effect could not be observed in the case of addition $F(1,272)=0.38$ $p=.536$ $\eta_p^2=.00$ but only in the case of subtraction $F(1,272)=87.44$ $p<.001$ $\eta_p^2=.24$ (see Figure 3).

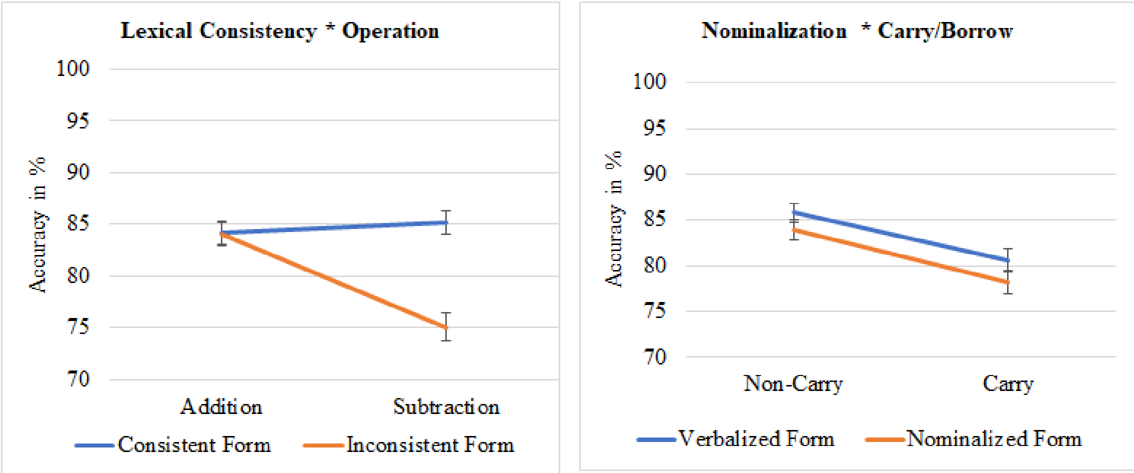

Figure 4S - 3 Interaction of Lexical consistency * Operation and Nominalization * Carry/borrow.

Summary. All four factors show a main effect. Furthermore, related factors interact, like operation * consistency – consistency effect is only present in case of subtraction. As well as, unrelated linguistics factors interact with other factors: nominalization * operation – nominalization effect is only present in case of subtraction. Moreover, unrelated mathematical factors over additively interact with other factors: carry/borrow * consistency, operation * carry/borrow. Finally, unrelated linguistics and unrelated mathematical factors do not interact with each other.

GENERAL DISCUSSION

SUMMARY OF THE RESULTS

The main goal of this dissertation was to broaden our understanding of the task characteristics (linguistic and arithmetic factors), the attributes of the individual (linguistic and arithmetic, but also general cognitive attributes) and the attributes of the environment (classroom and teaching attributes) that contribute to the difficulties of solving WPs. I also aimed to provide a better understanding of the process of problem-solving, for instance, by investigating if problem-solving processes interact and / or overlap with the processes playing a role in the building of a mental representation and to answer why even small arithmetic/numerical changes, for example, can affect performance and whether the origin of errors is solely due to simple representation errors. In order to explore the nature of the interaction between linguistic and arithmetic task characteristics I manipulated linguistic complexity to be independent of the arithmetic complexity underlying the WP in such a way that one factor was related to arithmetic / linguistic complexity and one factor was unrelated. Namely, the linguistic factor lexical consistency was related to arithmetic complexity while the linguistic factor nominalization was not. Analogously, the arithmetic factor operation was related to linguistic complexity, while the arithmetic factor carry / borrow was not. With the help of this manipulation it is possible to gain a deeper insight into problem-solving processes (i.e., whether the various factors influence the solving process to the same extent and whether the first reading phase and the calculation phase are separable from other problem-solving processes). Furthermore, how individuals with different characteristics and experience with different environmental factors (teaching quality) related to the task characteristics was evaluated.

In **Study 1** I proposed the theoretical process model, which integrates all of the components: task characteristics, individual characteristics and environmental factors. The results of the other studies (Study 2-4) partially support the model from Study 1.

In **Study 2** I tested adults with normal arithmetic and reading abilities. Carry / borrow, operation and nominalization factors showed a main effect on RTs. What is more, an interaction between consistency and operation affected RTs. Additionally, in this study no evidence was found for the overlapping of the initial reading phase and the calculation phase with other solution phases. The results show that errors do not necessarily originate from an incorrect mental representation but also from simple miscalculation, as no interaction with carry / borrow and the other factors was found, but ACC was influenced. In addition, domain-specific skills like literacy and numeracy skills influenced performance positively; however, a relationship between domain-general working memory performance and RT was not found. Finally, individuals that were more arithmetically capable had less difficulty with complex arithmetic conditions and better readers experienced less problems with more complex grammatical structures.

Based on my basic research with adults (Study 2) the second research goal was to explore the difficulties children have with WPs in **Study 3**. Operation and nominalization increased RT. For ACC, effects of consistency and operation were found only for sentences with the nominalized form (i.e., semantically more complex sentences) but not for easier sentences with the verbalized form. Additionally, increasing arithmetic difficulty shifted eye-movements towards not only the numbers but also the textual elements. Contrary to the adult study, interaction with the unrelated factors nominalization and the other factors were found, which might suggest that both reading and calculation phases partially overlap with the other solution phases. Finally, there was relationship between individual characteristics and sequential eye-movement behaviour. Children with better reading comprehension more often had non-sequential reading patterns. On the other hand, children who had lower visual working memory, reading skill and calculated slowly and less accurately preferred to read the WP word by word and did not regress to previously read sentences before fixating on the question.

The focus of **Study 4** was on which children solve some kinds of WPs with greater ACC and how this relates to other educational factors (teacher-student relationship and classroom management). Related factors interacted: consistency effect is only present in case of subtraction. As well as, unrelated linguistics factors interacted with other factors: operation effect was only present in semantically more complex sentences. Moreover, unrelated arithmetic factors over additively interacted with other factors: the consistency and the operation effects were especially pronounced in arithmetically more complex sentences (carry/borrow compared to non-carry/non-borrow). Fluid intelligence and reading and mathematical skills positively influenced the overall ACC. Cognitive activation and supportive climate were associated with higher WP solution ACC for problems with increased reading difficulty in the case of nominalization.

Generally, solving WPs depends on linguistic and arithmetic complexity, which are, as the interactions show, at least partially processed at common stages. Finally, beyond the characteristics of a WP, domain-specific inter-individual differences in both literacy and numeracy contribute to performance differences. An integrative summary of the results is discussed in the next paragraphs.

THEORETICAL PROCESS MODEL

Study 1 introduced a theoretical process model of WP solving (Daroczy et al. (2015). In this model three general aspects are distinguished for predicting individual WP performance. Task characteristics, individual characteristics and environmental factors. According to the model, WP difficulty comprises linguistic factors, arithmetic factors and their interaction. Individual characteristics can refer to linguistic and arithmetic capabilities and domain-general abilities such as individual working memory capacity. Task characteristics and individual characteristics influence WP performance both directly and through mediator variables (e.g., domain general attributes and specific solution strategies). For example, complex linguistic and arithmetic factors can increase cognitive load and the impact of this increased complexity may be over additive, especially when joint linguistic and arithmetic complexity exceeds an individual's cognitive load capacity. On the other hand, domain-general attributes are influenced by individual capability. Environmental factors influence individual characteristics, solution strategies and individual WP performance. (see Figure 3).

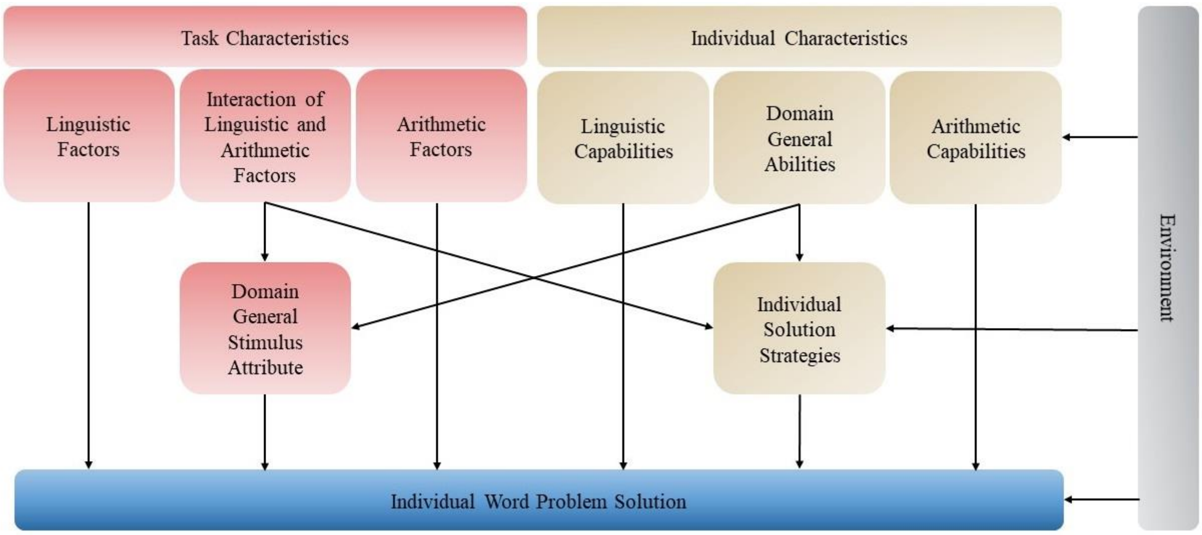

Figure 2 Theoretical process model (Study 1).

Based on empirical evidence from **Studies 2-4**, I would like to revisit the theoretical process model from Study 1 considering the following components:

- Task characteristics: The theoretical process model theorized that task characteristics 1) directly influence individual WP performance (it also aimed to look at the extent of influence of different linguistic and arithmetic factors) and 2) task characteristics influence individual WP performance through mediating factors. Additionally, this dissertation investigated how different task characteristics relate to the process of problem-solving and whether the initial reading phase and the calculation phase overlap with other stages of problem-solving.

- Individual characteristics: The model hypothesised that 1) individual characteristics directly influence ability to solve WPs and 2) individual characteristics indirectly influence ability to solve WPs.
- Environment: According to this model environment should have a direct influence on performance.

In addition, the theoretical process model extends existing WP solving models on several points. First, the model provides evidence for the problem model approach over the schema model; otherwise, I would not have observed interactions with unrelated factors if the solution relied solely on schemas. Additionally, in opposition to models that predicted that errors originated from problem misunderstanding, I demonstrated that miscalculation is a source of lower solution rates. With the distinction of related and unrelated factors, this dissertation demonstrated that problem-solving is not necessarily a linear process, but that initial reading processes and calculation processes might overlap with other problem-solving phases. With this distinction, the theoretical process model also extends previous WP solving models that did not fully answer why simple changes in linguistic and arithmetic difficulty alter performance. For instance, with the factor carry it was possible to investigate processes which are not just related to mathematical semantics (Gros et al., 2020). Finally, taking into consideration both individual characteristics and environmental factors, it is possible to understand the interrelation and interaction with the problem-solving process to a greater extent.

TASK CHARACTERISTICS - LINGUISTIC AND ARITHMETIC EFFECTS ON WPS

The theoretical process model of Study 1 theorized that task characteristics directly influence individual performance on WPs. This was confirmed in the empirical studies of this dissertation by overall performance (ACC), solution time and total attention to the text. Table 1 summarizes the results of the main effects and presents interactions. Interactions will be discussed in the section Results – Interactions. Moreover, dividing linguistic and arithmetic factors according to their relatedness could help to answer the question of whether or not all linguistic and arithmetic factors influence the process of WP solving to the same extent.

Table 1 *Summary of Results Across the Studies.*

	Adults Study 2 (*N*=23)		Children Study 3 (*N*=33)		Classroom Study 4 (*N*=273)
	RT	ACC	RT	ACC	ACC
Linguistic Factors					
Nominalization	N>V	--	N>V	*	*
Lexical Consistency	*	C>IC	--	*	*
Arithmetic factors					
Carry/Borrow	Ca>nC	nC>Ca	NI	NI	nC>Ca *
Operation	S>A *	--	S>A	*	*

N: Nominalized Form, V: Verbalized Form, IC: Inconsistent Form, C: Consistent Form, Ca: Carry / Borrow, nC: Non-Carry / Non-Borrow, S: Subtraction, A: Addition, NI: Not Included, * factor had an interaction with other factors which are elaborated in Figure 3, Figure 4, Figure 5.

Unrelated Factors. Regarding the unrelated linguistic factor ***nominalization***, the nominal form significantly increased solution time compared to the verbalized form for both children and adults. In line with previous literature (e.g., De Corte et al., 1985), this shows that even small changes to the text can lead to a difference in the solution process. Considering that the number of sentences, words and even characters was almost the same for all problems in this study, this finding provides support for the importance of linguistic factors unrelated to arithmetic in WP solving. Extra processing time was not paired with lower ACC for adults, only for children. This can be interpreted to mean that children are more error prone in the face of textual reading difficulty. Additionally, only children made more errors in subtraction compared to addition in sentences using nominalized form. This may be due to the higher cognitive load that children experience as compared to adults. Moreover, nominalized form lengthened attention and increased the number of re-readings of the total text as compared with verbalized form for children (Study 3). As the eye-movement measures for the total text represent the overall difficulty (Strohmaier et al., 2019), these results support the direct influence of the linguistic factor nominalization on the problem-solving process. Looking back at the text is especially important because it is associated with the second phase of problem-solving, and the results hint that there must be some reading comprehension after the initial reading phase in the case of children.

The unrelated arithmetic factor ***carry / borrow*** caused the greatest increase in solution time across the factors for adults. This is supported by previous findings that carry / borrow in subtraction increases arithmetic difficulty (Deschuyteneer et al., 2005). Although, I did not include carry / borrow as a factor in Study 3, I would still expect similar results – i.e., children need a longer period of time to solve WPs with carry / borrow than problems without. Also, in line with the study of Dresen et al (2020) the effect of the carry/borrowing procedure indicated an increase in task difficulty. Nevertheless, not only was RT influenced by the factor carry / borrow but also solution ACC. Moreover, both adults (Study 2) and children (Study 4) experienced more failures when a WP included a calculation with carry / borrow.

Related Factors. In case of the related arithmetic factor ***operation*** both adults (Study 2) and children (Study 3) needed more time for subtraction than for addition tasks. This finding is in line that solution strategies for subtraction differ depending on the complexity of the problem (Caviola, Mammarella, Pastore, & LeFevre, 2018). Dresen et al. (2020) has shown that understanding the actual arithmetic problem from the text specifically increases the difficulty of subtraction. This direct effect of operation on WP solving was also present in the eye-movement data (Study 3). This is consistent with, and extends similar findings, where arithmetic calculations were not embedded in a WP (Geary et al., 2012; Imbo et al., 2007). The results of Study 3 showed, that the numbers were affected by changes in the eye-movement measures (only by the factor operation and not by any of the linguistic factors) and there were more re-readings in the case of subtraction as compared to addition for both the whole problem and the text alone. As re-readings are associated with the phases after the initial reading phase this indicates that the text, and not just the numbers, is affected by the calculation. Therefore, the behavioural and eye-movement measures suggest an overlap of the calculation phase with other problem-solving phases in the case of children. The finding that irrelevant numerical information negatively influenced WP solving ACC, while linguistically irrelevant information did not influence the solution, also supports this finding (Pongsakdi et al., 2020). Pongsakdi et al. (2020) investigated a similar age group: 4th-6th grade children and provides a supportive argument for the results of this dissertation because they showed that arithmetic complexity might affect problem-solving more than textual difficulty.

For the related linguistic factor, **lexical consistency** there was an effect for children and adults on solution time only for a subset of WPs. For instance, the consistency effect was present for subtraction problems. Nevertheless, lexical consistency caused both groups to make more errors. Here, the findings of this dissertation are consistent with the results of Hegarty et al.

(1995) and Van der Schoot et al. (2009) and opposed to those of Verschaffel et al. (1992). Verschaffel et al. (1992) found a consistent effect for children but not for adults. Surprisingly, there was no main effect in case of lexical consistency on the eye-movements – i.e., increased attention to the overall text in the case of lexically inconsistent sentences was comparable to that for lexically consistent sentences – in the current study. However, further analysis showed that the consistency effect was more pronounced and localized to the keywords, which is in line with Verschaffel et al. (1992) who hypothesised that the consistency effect manifests in the second sentence containing the keyword. This could mean that in the case of the consistency effect more sophisticated processes are involved, which focus not on the overall text, but only on part of it. Another explanation is that the consistency effect is not always found. For instance, Verschaffel et al. (1992) and Thevenot (2010) argued that mathematical keywords do not necessarily play a key role in WP solving.

In sum, these results show that in some way all factors contribute to the overall difficulty of a WP, but not to the same extent. This means that the main effects found in the studies of the dissertation support the direct influence of stimulus attributes on WP performance as in the proposed theoretical process model. All four factors had an effect on the solution time and / or ACC for both children and adults. Subtraction and nominalized form increased processing time for both groups as did carry / borrow for adults (not measured with children). The related linguistic factor lexical consistency did not influence the processing time but affected ACC for both children and adults. This means that although participants did not take longer in general for lexically inconsistent sentences, they more often solved the problems incorrectly. This finding partially contradicts some of the previous literature where WP solvers made not only more errors but took longer to solve inconsistent, as compared to consistent, WPs (e.g., Pape, 2003). An explanation for the results presented in this dissertation might be that there could be a speed ACC trade-off. Namely, participants who make more errors on the inconsistent problem quickly pick the numbers and the keywords, which would also explain why these participants are more error prone. Additionally, children made more errors when faced with problems containing operation and nominalization and the eye-movement data reflected these results. Namely, items that were more complex reinforced longer, and more numerous eye-movements. I showed that with increasing linguistic difficulty eye-movements are shifted towards the textual elements and with increasing difficulty of the arithmetic information eye-movements are shifted towards numerical elements.

TASK CHARACTERISTICS - INTERACTIONS

In the model theorized in Study 1, it was hypothesized that WP difficulty comprises not only the linguistic complexity of the text and the arithmetic complexity of the arithmetic problem but also their interaction as some attributes are processed at common stages. Additionally, in the already existing problem-solving models, it is not clear whether problem-solving phases are fully separable or not. Manipulating related and unrelated factors at the same time might allow for the elaboration of how the presence or absence of an interaction between the factors could support existing models, for example, the absence of the interaction between unrelated and related factors could support overlapping processing stages.

Interaction of related factors. I expected related factors (***consistency * operation***) in WPs to interact with each other because related factors are strongly connected to the building of problem models; therefore, this result was expected from both the sequential, because consistency and operation can be linked to the problem, and cyclic models. The sequential model would expect this interaction because consistency and operation should affect the problem model phases. Indeed, the interaction was stable across the studies. As expected, operation interacted with lexical consistency for adults in the case of solution time (see Figure 3), as well as for children. Namely, the consistency effect was only present in the case of subtraction (i.e., the more difficult arithmetic operation which led to an extra demand on problem-solving skills for adults). Additionally, the consistency effect was present in the case of subtraction than for addition in the case of ACC (see Figure 3) in Study 4 which is in line with the results of Study 2.

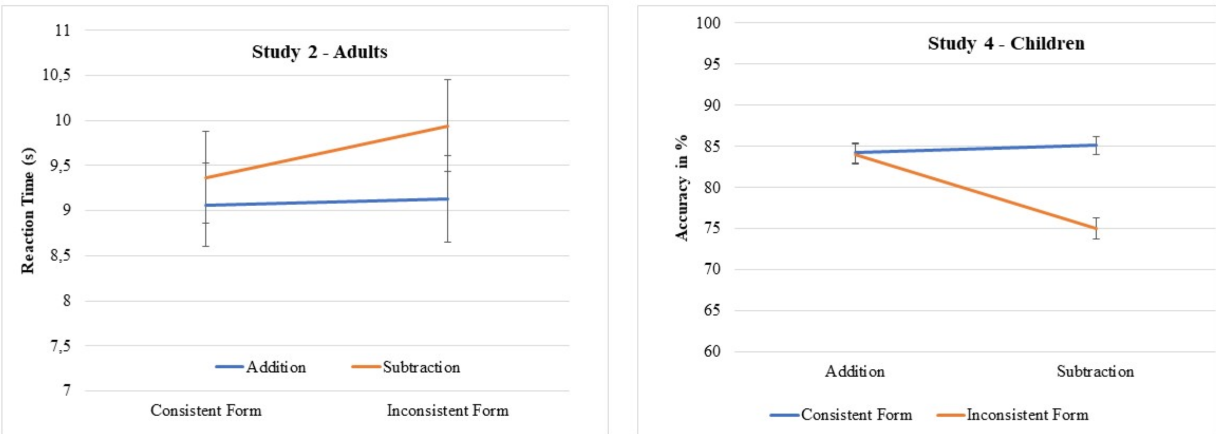

Figure 3 Interaction of the Related Factors Operation * Lexical consistency (Study 1 - RT) and Children (Study 4 - ACC).

Finally, in the eye-movement analysis, interactions were observed with specific parts of the WPs. For instance, in arithmetically complex conditions (i.e. subtraction) participants looked longer at the textual parts of the WP. Numerical areas, however, seemed not to be affected by linguistic difficulty.

Interaction of unrelated factors. As stated above, the presence of an interaction between unrelated and related factors might provide a deeper insight into if the initial reading phase and the calculation phase interact with other problem-solving phases. In Study 2 there was no interaction between related and unrelated factors but in the Study 3 and Study 4 several interactions were found between related and unrelated factors which will be elaborated in the next sections.

Initial reading phase and other phases of problem-solving. An interaction between nominalization and the other factors suggests that the reading comprehension phrase interacts with other problem-solving phases. In Study 2, the unrelated linguistic factor (nominalization) did not show an interaction with other factors. The results favour the propositional model over the non-sequential and cyclic models for adults. Therefore, a possible interpretation is that the non-related linguistic factor might affect only the initial reading phase for adults.

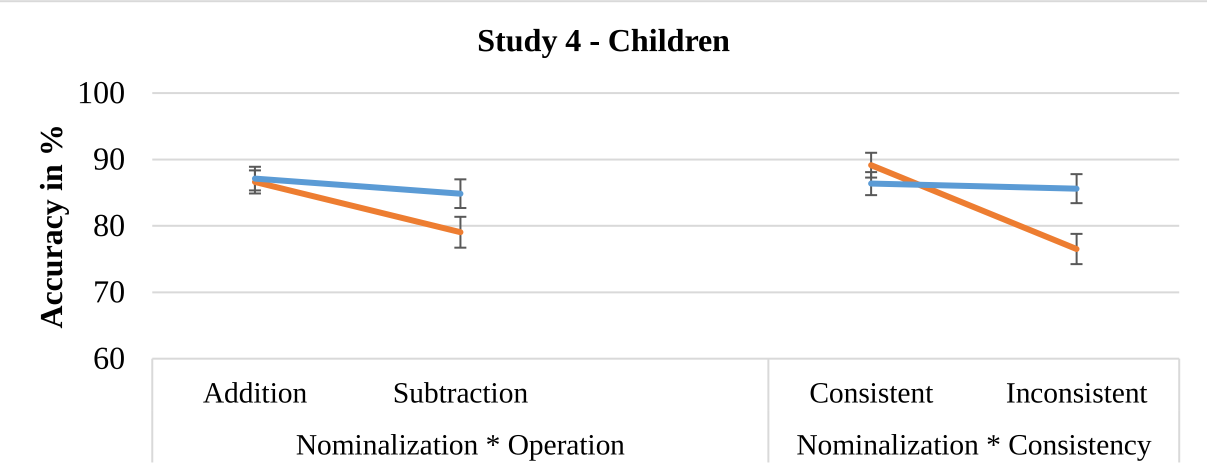

Figure 4 Interaction of Nominalization * Operation and Nominalization * Lexical consistency (Study 4).

However, children seem to be more sensitive to language difficulties: in both Study 3 and Study 4 children made more errors when solving subtraction problems, as compared to addition problems, when the problems contained nominalized sentences, but not when they contained verbalized sentences. This can be interpreted to mean that it is harder for children (who are by default less experienced than adults) to construct a good model when the language is complicated, and nominalized form makes the transformation into the model harder. This could be interpreted as having purely to do with text comprehension and synthesis, but it is important to note that in both verbalized and nominalized sentences the underlying mathematical structure is the same. Additionally, the consistency effect was present only for problems with nominalized sentences and not for those with verbalized sentences (see Figure 4), which means that in texts which are easy to read participants are equally capable of solving lexically inconsistent sentences where creating the correct the mental representation is more difficult. The following observation can be made: As soon as the language gets more difficult (despite the identical underlying mathematical structure) participants are unable to perform the extra processing needed for the inconsistent cases. This may be due to one reason: nominalized form blocks working memory capacities and causes a cognitive overload in combination with the lexically inconsistent cases. This could have direct applicability as the ease of constructing the mental representation could be influenced by differences in text formulations. This also means that the additional processing difficulty of the related factors emerged in the linguistically more demanding WPs. This finding would also speak to overlapping processes in the initial reading phase and the comprehension phase for children. The results from the eye-tracking data also speak to an interaction between these problem-solving processes because nominalization affected how often children re-read the text, which is associated with the phases after that of the first reading. It is likely that as adults have greater reading comprehension skills, therefore the factor nominalization does not affect the working memory resources to the same extent as in the case of children.

In sum, the results indicate that for adults there is no overlapping processing stage in the initial reading phase and other phases, but for children the reading comprehension phase interacts with the problem model-building phase.

The calculation phase and other phases of problem-solving. The presence of an interaction between the non-related arithmetic factors carry / borrow and the other factors may answer whether or not the computational process also affects the problem model building phase. There were no interactions between carry / borrow and other linguistic or arithmetic factors in the case of adults (Study 2). This is again in favour of the propositional theory, which sees the calculation phase and other problem-solving phases as distinct and is in line with Rabinowitz and Wooley (1995) who found no interaction of either RT or ACC with the factors problem size or carry / borrow or other factors. This means that in this adult study I found no evidence that number difficulty affects another stage of problem-solving besides the calculation phase. However, errors were increased for adults. This may be interpreted to mean that errors do not necessarily originate from the formation of an incorrect problem model, as usually expected (e.g., Lewis & Mayer, 1987), but simply from miscalculation. This could be also interpreted – opposed to (Gros et al., 2020) – that the first problem model is more fragile, but not necessarily wrong.

Nevertheless, for children (Study 4) carry / borrow interacts with the factors operation and consistency in the case of ACC (see Figure 5). Namely, the consistency and operation effects were especially pronounced in the case of carry / borrow as compared to non-carry / non-borrow. This means on the one hand, both consistency and operation effects grow larger when tasks involve complicated calculations and that it is even more difficult to build the correct problem model. On the other hand, this can be interpreted to mean that for children the factor carry / borrow cannot be separated from the phase where the problem model is built. In sum, the results indicate that numerical difficulty is independent from the construction of a problem model for adults, but for children numerical information in the text might play a bigger role in WP processing, as previously assumed.

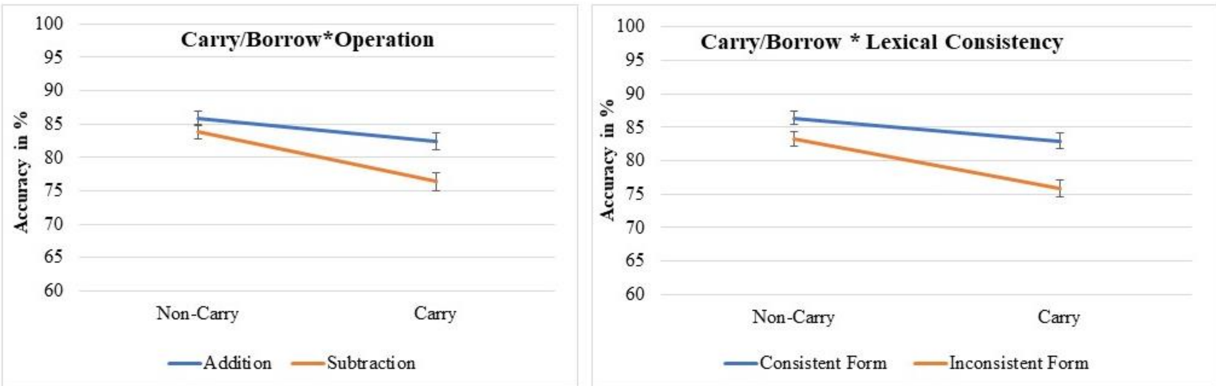

Figure 5 Interaction of Carry/borrow * Operation and Carry/borrow * Lexical consistency (Study 4).

Finally, we should have a short look at the interaction between **carry / borrow** and **nominalization.** This interaction hints towards a common processing stage between the initial reading phase and the calculation phase. None of my studies have shown this interaction, even the results from Study 3 show that simple linguistic manipulation does not increase attention paid to the numbers, which might indicate, that the initial reading phase and the calculation phase are two distinct domains.

To summarize the results, in the Study 2 there was no interaction between related and unrelated factors. However, in the Study 2 based on the model of Daroczy et al. (2015) an interaction would be expected in the case of factors where the cognitive load increases. Perhaps cognitive load increased sufficiently only for children, and not adults, which created a bottleneck in the joint processing stage, which would explain the differences between children and adults. The lack of interaction between unrelated linguistic and arithmetic factors could also mean that it is not correct to assume a limited load domain-general stage model for adults as described in Study 1. Nevertheless, in the child studies several interactions were found between related and unrelated factors, which hints towards common processing stages for children. However, adults and children should not necessary be the same in that respect because adults also have more proven strategies they could apply without needing to build different problem models. In addition, the absence of an interaction between related and unrelated factors is more consistent with the propositional theory than with the cyclic model for adults. This means that for adults I have found no evidence that number difficulty affects another stage of problem-solving besides the calculation phase. This implies that, for adults, the initial reading phase, the calculation phase and the building of the problem model phase can be viewed as distinct, non-overlapping stages of problem-solving. However, the interaction with the related factor could also be seen as a supportive argument for the cyclic model –for the overlapping of the calculation and problem model phase. Nevertheless, we cannot be fully sure how clearly the related and unrelated factors can be separated from the problem-solving phase. In my view, the interaction between the related factors is not clear proof for the existence of overlapping problem-solving phases.

It seems from the results of Study 3 that for children the initial text comprehension phase interacts with the other phases of problem-solving. This is supported by De Corte et al. (1990) who questioned the sequential and linear character of child WP solving particularly with respect to more complex problem types. However, contrary to the case with children, these solution phases might be more clearly distinguished for adults. This means that adults and children differ in their problem-solving processes and strategies. This is supported by the fact that children learn with time to distinguish between relevant and irrelevant information (Van der Schoot et al., 2009), also the eye-movements of preschool children differ somewhat from those of adults as they have more frequent and smaller saccades (Hegarty et al., 1995; Rayner, 1998). These differences hint towards developmental change, meaning that adults approach WPs with more automatized processes than children do. However, it might be that this is only applicable to easier WPs, as the WPs used in the Study 2 were rather easy for the adult participants.

INDIVIDUAL CHARACTERISTICS

Direct influence. The model in Study 1 also theorized that individual characteristics influence WP performance both directly and indirectly, and that the cognitive load should be higher for an individual with lower linguistic and numerical abilities. This turned out to be partially true. In Study 2 (adults) numeracy and literacy skill negatively correlated with RT. In addition, in Study 4 (children) higher reading speed, reading comprehension skill, intelligence and mathematical skill were associated better solution ACC. The hypothesis that arithmetic complexity should affect participants with better mathematical skills less than participants with lower mathematical skills and the linguistic complexities should affect participants with better reading skills less than participants with lower reading skills were partially confirmed. For instance, in Study 2, individuals that were more arithmetically capable had less difficulty with complex arithmetic conditions. Similarly, better readers experienced less problems with more complex grammatical structures. In Study 4, better reading comprehension and higher intelligence accounted for greater ACC when dealing with nominalized sentences. This indicates that good reading skill helps a participant solve the problem, especially when the linguistic part of the WP gets more difficult. This means that both adults and children with better reading skills are less affected by linguistic complexity. Generally, these results are in line with previous results. For example, Fuchs, Geary, et al. (2016) and M. O. Martin and Mullis (2013) argue for a strong role of language comprehension skills in WP solving. Additionally, the finding that increasing mathematical skill accounted for better ACC (children) and decreased RTs (adults) is in line with previous research; namely, that mathematical skill is a necessary foundation for solving WPs (Fuchs et al., 2015). As expected, intelligence accounted for better WP solution ACC in the case of children (Study 4). This is in line with previous research (Mori & Okamoto, 2017; Swanson & Beebe-Frankenberger, 2004) that suggests a direct relationship with WP solving. Highly intelligent children made fewer errors when faced with sentences containing nominalization. Additionally, in the same study, highly intelligent children make more calculation errors under carry / borrow conditions. This result was surprising, because I expected a smaller carry effect for highly intelligent children because they usually perform better on more complex tasks than children of normal intelligence (Xin & Zhang, 2009). The reason could be a speed-ACC trade-off; highly intelligent children solved the problems quickly but at the cost of making more mistakes. Indeed, there was a weak positive correlation between the number of fully solved items and the number of errors (r (359) = .16, p = .003.) in the total sample. In the carry / borrow effect analysis I included only those participants who had answered the first 32 WPs, not only the first 16 WPs. This criterion reduced my sample size by 3/4 (i.e., 100 children were excluded), so theoretically the fastest children were included in this analysis. Finally, contrary to the findings of the OECD (2013), which showed that more socioeconomically advantaged students have higher scores in mathematics, I found no effect of socioeconomic status in Study 4. One explanation could be that the socioeconomic question did not differentiate enough between students, as the entire group was in a gymnasium school with comparable socioeconomic backgrounds. Although I did not find an impact of socioeconomic status on general WP performance, higher socioeconomic status led to better ACC on the linguistically more demanding tasks. The reason might be that children with high socioeconomic status are more familiar with the use of nominalization. Indeed, academic language, which is especially dense in nominalization is considered to be especially difficult for students from low socioeconomic status families (e.g., Bailey, Butler, LaFramenta, & Ong, 2004). Furthermore, long and complex words are associated with difficulties for students from low-socioeconomic status families (Heppt, Haag, Böhme, & Stanat, 2015)

Indirect Influence over mediator variables. Working memory had no effect either for adults or for children. This is surprising because many past studies (Adams & Hitch, 1997; Furst & Hitch, 2000; Passolunghi & Siegel, 2001; Swanson, 2004) have shown a strong influence of working memory on mathematical performance and only a few studies did not find working memory to be a significant predictor (Fuchs et al., 2015). The explanation could be that working memory load depends on WP difficulty. For instance, according to Peng et al. (2016), multistep mathematical tasks that require the calculation and maintenance of intermediate results are hypothesized to use more working memory resources than mathematical tasks consisting of fewer steps. In the current studies, I used one-step WPs, which would indicate lower cognitive load. Finally, fluency in WP solving impacts the extent to which individuals rely on working memory (Spencer, Fuchs, & Fuchs, 2020). In addition, individual factors might also affect one another. For example, for the nominalization effect, mathematical competences were not a significant predictor, but in Study 4 there was a correlation between mathematical skill, language competences and motivation skills, which indicates a relationship between these factors and a potential indirectly influential role. Mathematical skill is positively correlated with language competences, which influences the nominalization effect directly. This is also supported by Swanson (2004), who found that arithmetic skill mediated the relation to WP solving through reading and intelligence.

In general, the data are consistent with the idea that arithmetically complex WPs are particularly difficult for students with poor arithmetic capabilities and linguistically complex WPs are particularly difficult for students with poor reading skills. In sum, individual characteristics play an important role in WP solving. However, less is known about the relationship between individual characteristics and individual strategies.

INDIVIDUAL STRATEGIES

The second mediator variable of the theoretical process model referred to specific solution strategies. This was not the focus of this thesis; however, it was investigated in the studies and preliminary results from the eye-movement data indicate that there are several strategies. In this dissertation I conducted a preliminary analysis to investigate the relationship between individual abilities and eye-movement patterns. For this, I categorized participant reading patterns into systematic sequential patterns (every sentence is read word by word) and non-sequential patterns (e.g., the question is read first, or already read parts of the text are read again before the entire problem is read). However, similarly to the study of De Corte and Verschaffel (1986) several other patterns could be observed: WPs can be read sequentially, or certain parts can be re-read before arriving at the end of the first reading

I observed individual differences in reading strategies. In addition, children with better reading comprehension more often had non-sequential reading patterns and read the question before reading the first, second, or third sentence. This is supported by Thevenot et al. (2007) and by Cook and Rieser (2005) who found that highly capable students used question-guided strategies more often. Also De Corte and Verschaffel (1986) observed that most rereading was done children in the high-ability group. On the other hand, children who had lower visual working memory and reading skill read the sentences sequentially word by word without looking back to previously read sentences before fixating on the question. Although there was no main effect of working memory on various types of WPs in the current studies, it seems that it affects specific strategies. In addition, children who had better mathematical skills had a different reading pattern. Namely, children who calculated rather slowly and less accurately read the sentences sequentially word by word and did not regress to previously read sentences before fixating on the question, or they did not start with any other sentence than the first one. Thevenot et al. (2007) observed better performance when the question was presented prior to the text, especially in the case of children who had poor mathematical skills. This also could mean that for children with poor mathematical skills the process of calculation is even more closely integrated into other stages of the WP solving processes than for individuals with better mathematical skills.

In sum, the strategies might depend on the formulation of the text and different participants use different reading strategies. The preliminary results of this dissertation show that individual abilities influence eye-movement patterns and we should question the dominance of systematic reading patterns. Whether people read in a strictly systematic manner or use other non-sequential strategies depends on age and reading skill. Individual abilities influence eye-movement patterns and the results of this dissertation support the proposed theoretical process model.

ENVIRONMENT

The theoretical process model proposed that the environment has a direct influence on the solving of WPs. This assumption was supported by the results because students from classrooms with higher cognitive activation performed better. Although, in Study 1 the model proposed no connection between the task characteristics and environmental factors, the results of Study 4 show that environmental factors are probably linked to task characteristics more strongly than previously assumed. In Study 4, I investigated how teaching quality (classroom management, cognitive activation and supportive climate) affects WP solution ACC while distinguishing between linguistically and arithmetically simple and complex WPs. Environmental factors affected performance when children were solving linguistically demanding items, namely a more supportive climate and higher cognitive activation reduced the nominalization effect (i.e., children achieved higher ACC on WPs in the nominalized form as compared to the verbalized form). The results indicate that students from classrooms with good cognitive activation and a supportive climate can solve linguistically more demanding tasks with higher ACC. However, the effects were quite small. This is important because high instructional quality might account for better mathematical achievement (Adelson et al., 2015).

Contrary to the expectations, there was no difference between the classes on the consistency, operation and carry effects. This is surprising because higher cognitive activation should reinforce the problem model strategy over the direct translation strategy (Hegarty et al., 1992). One possible explanation is that the consistency effect is indirectly influenced by teaching quality measures. This is supported by the fact that the consistency effect positively correlates, for instance, with the motivational scale, meaning higher individual importance generally results in better ACC on inconsistent tasks. Last, but not least, I also expected that better classroom management to lead to better ACC on more difficult tasks (based on the results from Fauth et al. (2014) who found classroom management to have a direct role in student performance) but classroom management was not a predictor in any of the models. Additionally, the results of the dissertation show that aggregated classroom management values correlated with total performance, indicating a strong relationship between WP performance and the variance within different classes.

SUMMARY

In sum, I suggest not only that linguistic and arithmetic factors should be differentiated but also that there should be a differentiation between related and unrelated factors in WP solving processes. The distinction between related and unrelated factors is the essential new aspect of the revised theoretical process model (see Figure 6). This is important because as related factors might generally rely more strongly on working memory capacities, and be more strongly connected to the underlying problem model, than unrelated factors. Additionally, the existing problem-solving models do not explain the differences in problem-solving ACC or unrelated factors. For instance, in the case of unrelated factors, according to the propositional model, underlying problem models should be unaffected. For example, changes in carry / borrow could not be explained by the semantic model. This dissertation demonstrated that different task characteristics have different influences on the problem-solving process in different groups. In other words, the process of problem-solving is influenced by the task characteristics and is also affected by environmental factors.

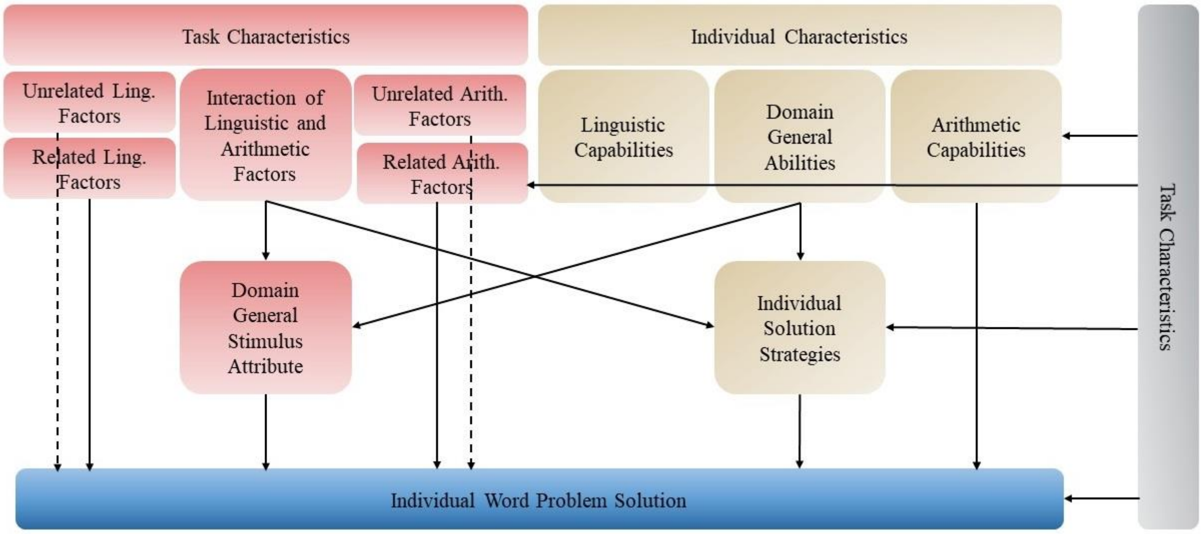

Figure 6 Revised Theoretical Process Model.

Furthermore, not all of the factors (especially the unrelated factors) have influence on WP performance directly. There are factors which influence the cognitive load in combination with individual characteristics. For example, for children, the problem-solving process changes not only when the mathematical structure of the text is different, but also when the text is formulated differently and has another mathematical structure. As described in the introduction, none of the existing models of the WP-solving process provide a clear answer as to why simple linguistic or numerical modifications of the problem that do not change the underlying mathematical structure affect the solution process. The models assume that the initial representation is incorrect when the processing time is longer, and that a higher error rate is due to simple linguistic or numerical modifications. For instance, the SECO model from Gros et al. (2020) explained that in such cases individuals have to recode this initial representation into an alternative representation. Contrary to that, in the case of a simple numerical modification like carry/borrow vs. non-carry/non-borrow the initial representation should be equivalent, and not influence other stages of problem-solving processing. But the results of the carry effect in this dissertation suggests that the mental representation gets more complicated owing to the increasing difficulty of the numerical processing. One explanation could be that the processes of calculation and the problem model do not necessarily overlap but rely on the same cognitive resources and increase the cognitive load. This is supported by the previous finding that working memory demands are increased by carry/borrow operations (Imbo et al., 2007). Nevertheless, the interactions between carry/borrow and other factors found in Study 4 suggest that the mental representation is affected by the carry/borrow factor. This means that numerical processing should happen earlier than assumed and the initial representation is ready. This is only possible if the processes of calculation and mental representation do indeed overlap and have a common processing stage. This implies that the first problem model is probably more fragile, but not necessarily wrong, as suggested by the results of Study 2. Thus, re-reading the text serves to strengthen the problem model. This is possible if the initial reading and calculation phases overlap with the other processes and would provide a more flexible way of dealing with cognitive resources as well as explain those cases involving unrelated factors.

In addition, the results indicate that the interaction depends on the experience level as there is a difference between adults and children. Related factors interact in both adults and children, but the unrelated ones seem to depend on the level of expertise – i.e., there are more interactions present in the case of children. Moreover, the cyclic or sequential processing also depends on the expertise level of the problem solver, because for children the overlap is stronger – both the initial reading phase and the calculation phase interact with the phases where the problem model is built — but for adults it is almost not present. This indicates that the process of solving WPs might be different for adults and children; the problem-solving method may change depending on year of education and experience. This means that a single problem-solving model might not be enough and thus would result in two separate models for children and adults. Therefore, the model considers these two aspects: both for children and adults. The interactions which are more likely exhibited by children, not adults, are marked by striped lines in Figure 6. The hypothesis that adults and children differ in their problem-solving models is supported by Cummins et al. (1988) who suggested that younger children have difficulties at the comprehension stage while older students have difficulty with the representation of the situation model described in a text. On the other hand, solution strategies – whether people read in a strictly systematic manner or use non-sequential strategies – depends both on the formulation of the text and on the individual's capabilities. Finally, the theoretical process model in Study 1 proposed no connection between the task characteristics and environmental factors. However, the results of Study 4 show that environmental factors are linked to task characteristics more strongly than previously assumed (see Figure 6).

LIMITATIONS & FUTURE DIRECTIONS

The current dissertation showed the importance of differentiating between factors in the processes of problem-solving. The theoretical process model suggested a connection to individual strategies – an aspect not fully covered in this dissertation. One potential future undertaking is to investigate how strategies change based on difficulty level. For example, for easier problems, fast processes – e.g., fact retrieval or the simple keyword searching strategy might be sufficient. For easier problems it is also possible that problem-solving processes do not overlap. For more difficult problems, these fast processes might mislead problem solvers. It is possible that for more difficult problems heuristic strategies are more effective, and problem-solving phases are harder to distinguish for both adults and children. Furthermore, I have shown that not only do linguistic and arithmetic task characteristics influence the solution rate on the behavioural level but also on the level of eye-movement (i.e., different patterns), and interaction can be observed for a limited set of WPs (i.e., change WPs). However, there are several potential factors that could also affect the process of WP solving.

TASK CHARACTERISTICS

Other linguistic and arithmetic factors that are potential candidates for systematic study include number sense (Dehaene, 2001), the presented order of the numbers, and the influence of these features on solution ACC, paired with cognitive ability and item difficulty. Specifically, lexically inconsistent language in subtraction problems and a smaller number in the first position may increase RT. Furthermore, in subtraction WPs the subtrahend can be represented before the minuend, which is inconsistent with the order of the numbers in the solution. Subtraction tasks where the smaller number is presented first can be solved with multiple strategies (De Corte & Verschaffel, 1987; Torbeyns et al., 2009). Alternatively, factors other than those investigated in this dissertation might affect the solution process. For example, in a study by Bagnoud et al. (2018), brain activity differed during WP solving depending on if the problem was about discrete quantities (e.g. apples), or continuous quantities (e.g. sand). Bassok et al. (2008) showed that semantically related words prime addition and lead to slower RTs. Such semantically related factors may influence other processing stages differently than, for instance, nominalization. It is also unclear if incorrect solutions result from the construction of an inappropriate problem model, or if the language potentially triggers the wrong calculation. Therefore, I suggest that further elements of text characteristics that cause difficulty for various individuals must be identified in a systematic manner.

Individual Characteristics

It must be distinguished whether the incorrect solution results from the individual not being able to create a problem representation, or if it results from (despite the correct problem representation and otherwise good mathematic skills) a calculation error due to increased cognitive load due to the interaction between linguistic and arithmetic factors. In both cases, the result is the same, namely an incorrect solution, but each case requires a different intervention. Therefore, I suggest a differential-psychological approach to WP research. Different students may have problems with different types of WPs. Only with such differentiation on an item level (as regards linguistic and arithmetic complexity and their interrelation) and on an individual level (as regards linguistic and arithmetic skills and general cognitive abilities) will it be possible to understand why a child has their individual difficulties with certain types of WPs (Daroczy et al., 2020).

Finally, individual strategies matter. Although, there is a difference reported between students of high and low ability (Hegarty et al., 1995; Van der Schoot et al., 2009) in terms of looking back at numbers and text in WPs there is still little known about cognitive abilities and the eye-movement patterns in this context. Therefore, further research is needed to explore and categorize reading patterns along with integrating reading models and specific reading characteristics.

EDUCATION

It is important to note that the WP is a didactical construct that serves purposes in mathematical education from simple exercises on basic operations (Greer, 1997) to modelling (Verschaffel et al., 1997). WPs are used very broadly and range from the presentation of a simple calculation to complex problem-solving. It has been shown that the role of the WP varies across curricula and teachers (Baumert et al., 2010; Nortvedt et al., 2016). However, there is not much known about how exactly WPs influence behaviours and how WPs can be targeted. For instance, if teachers would like to demonstrate how a mathematical operation works, at some stage it would be unnecessary to increase the linguistic difficulty as it would also affect the calculation. This information should be specified and considered for future interventions and curricula development.

METHODS

The assumption is that in the case of incorrect solution students did not create a proper problem model (Greer, 1997; Hegarty et al., 1995), and instead used simple keyword searching strategies. However, this assumption is very hard to investigate on a behavioural level. For instance, to understand underlying models and the difference between strategy selection (e.g., the differences between the direct translation strategy, problem model strategy, or the connection of linguistic and arithmetic processes) measures would be needed that capture processes at the moment of their occurrence. Eye-movement word level analysis is such an option but it does not provide information about ongoing processes, like a scan path analysis. Word level analysis does not necessarily contain all the information about processing because many words can be skipped (Rayner, 1998). For fixation and regression time it is hard to dissociate the competence level of the subjects, as difficulty increases these measures, but it has also been shown that the higher the expertise of the specific subject the fewer the eye fixations needed to process information in specific zones. Therefore, applying different methods could provide a valuable extension to the current findings. For instance, there are indeed few studies on the neural correlates of WPs due to the complexity of the problems (e.g., Zhou et al., 2018). An understanding of the cognitive and neural mechanisms underlying arithmetical learning could contribute to our understanding of general cognitive development. Very little is known about the behavioural and neurological mechanisms that drive arithmetic WP solving.

IMPLICATIONS

The educational implications of the results of this dissertation are relevant to research. The direct implication of this dissertation is that mathematical and language requirements and resources need to be considered together in WPs. Recent findings argue for the importance of also training language in mathematics classes (Fuchs et al., 2018; Kyttälä & Björn, 2014) and suggest a stronger focus on training mathematical literacy and reading skills (i.e., Fuchs et al., 2018) not calculation skills. Nevertheless, the results of this dissertation suggest that it might be an incomplete view to mainly focus on the textual difficulty of WPs. Therefore, I would like to suggest that the involvement of arithmetic factors should not be neglected, even if children are able to make equivalent mathematical calculations in less resource-demanding tasks. This should be paired with high quality instruction (please note, Nortvedt et al. (2016) showed that only twelve percent of the teachers studied offered high quality instruction in both reading and mathematics) as this dissertation also provides evidence that environment influences WP performance. The fact that cognitive activation and supportive climate reduced the nominalization effect means that, for instance, positive and constructive teacher feedback, a positive approach to student errors and misconceptions, the integration of challenging tasks and the exploration of concepts and ideas can help to achieve a higher solution rate, especially for linguistically complex items.

The results also indicate that for some types of WPs, mathematical difficulty might play a particular role, while for others linguistic difficulty is more important. I suggest to confront children with a particular type of WP, namely, that from which they can learn the most. Instruction in text or mathematics may even need to be individually adapted. Interactive learning environments have shown promising results so far (Yamamoto et al., 2014). Also, in diagnostic assessments both the item characteristics (i.e., linguistic and arithmetic complexity) and the individual characteristics (i.e., general linguistic and arithmetic skill), as well as the environment, should be considered.

REFERENCES

Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). *Language Accommodations for English Language Learners in Large-Scale Assessments: Bilingual Dictionaries and Linguistic Modification. CSE Report 666*. Retrieved from https://files.eric.ed.gov/fulltext/ED518681.pdf

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied measurement in education, 14*(3), 219-234.

Abedi, J., Lord, C., & Plummer, J. R. (1997). *Final report of language background as a variable in NAEP mathematics performance*: National Center for Research on Evaluation, Standards, and Student Testing ….

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological bulletin, 131*(1), 30.

Adams, J. W., & Hitch, G. J. (1997). Working memory and children's mental addition. *Journal of experimental child psychology, 67*(1), 21-38.

Adelson, J. L., Dickinson, E. R., & Cunningham, B. C. (2015). Differences in the reading–mathematics relationship: A multi-grade, multi-year statewide examination. *Learning and Individual Differences, 43*, 118-123.

Aiken Jr, L. R. (1971). Verbal factors and mathematics learning: A review of research. *Journal for Research in Mathematics Education, 2*(4), 304-313.

Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 21*(9), 1201-1216.

Artemenko, C., Moeller, K., Huber, S., & Klein, E. (2015). Differential influences of unilateral tDCS over the intraparietal cortex on numerical cognition. *Frontiers in human neuroscience, 9*, 1-8.

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.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*(1), 75-106.

Ashkenazi, S., Rubinsten, O., & Henik, A. (2009). Attention, automaticity, and developmental dyscalculia. *Neuropsychology, 23*(4), 535.

Bagnoud, J., Burra, N., Castel, C., Oakhill, J., & Thevenot, C. (2018). Arithmetic word problems describing discrete quantities: EEG evidence for the construction of a situation model. *Acta psychologica, 190*(1), 116-121.

Bailey, A. L., Butler, F. A., LaFramenta, C., & Ong, C. (2004). Towards the characterization of academic language in upper elementary science classrooms. *Los Angeles*.

Baratta, A. M. (2010). Nominalization development across an undergraduate academic degree program. *Journal of Pragmatics, 42*(4), 1017-1036.

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.

Bassok, M. (2001). Semantic alignments in mathematical word problems. *The analogical mind: Perspectives from cognitive science*, 401-433.

Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology, 35*(2), 99-134.

Bassok, M., Pedigo, S. F., & Oskarsson, A. T. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, 34*(2), 343-352.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133-180.

Berends, I. E., & Van Lieshout, E. C. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction, 19*(4), 345-353.

Bergqvist, E., Dyrvold, A., & Österholm, M. (2012, January). *Relating vocabulary in mathematical tasks to aspects of reading and solving.* Paper presented at the MADIF 8, The Eighth Swedish Mathematics Education Research Seminar, Umeå.

Bergqvist, E., & Österholm, M. (2010, January). *A theoretical model of the connection between the process of reading and the process of solving mathematical tasks.* Paper presented at the MADIF 7, The Seventh Swedish Mathematics Education Research Seminar, Stockholm.

Boonen, A. J., de Koning, B. B., Jolles, J., & van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in psychology, 7*, 191.

Boonen, A. J., van der Schoot, M., van Wesel, F., de Vries, M. H., & Jolles, J. (2013). What underlies successful word problem solving? A path analysis in sixth grade students. *Contemporary Educational Psychology, 38*(3), 271-279.

Boonen, A. J., van Wesel, F., Jolles, J., & van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research, 68*, 15-26.

Boote, S. K., & Boote, D. N. (2018). ABC problem in elementary mathematics education: Arithmetic before comprehension. *Journal of Mathematics Teacher Education, 21*(2), 99-122.

Brissiaud, R., & Sander, E. (2010). Arithmetic word problem solving: A situation strategy first framework. *Developmental Science, 13*(1), 92-107.

Brock, L. L., Nishida, T. K., Chiong, C., Grimm, K. J., & Rimm-Kaufman, S. E. (2008). Children's perceptions of the classroom environment and social and academic performance: A longitudinal analysis of the contribution of the Responsive Classroom approach. *Journal of School Psychology, 46*(2), 129-149.

Brysbaert, M., Fias, W., & Noël, M.-P. (1998). The Whorfian hypothesis and numerical cognition: istwenty-four'processed in the same way asfour-and-twenty'? *Cognition, 66*(1), 51-77.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In F. K. I. Vedder (Ed.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 23-46). Amsterdam: John Benjamins.

Burton, L., & Morgan, C. (2000). Mathematicians writing. *Journal for Research in Mathematics Education, 31*(4), 429-453.

Butler, F. A., Bailey, A. L., Stevens, R., Huang, B., & Lord, C. (2004). *Academic English in Fifth-Grade Mathematics, Science, and Social Studies Textbooks. CSE Report 642*. Retrieved from https://files.eric.ed.gov/fulltext/ED483409.pdf

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition, 130*(3), 278-288.

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

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 179-202.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology, 3*(1), 21-29.

Carraher, T. N., Schliemann, A. D., & Carraher, D. W. (1988). Mathematical concepts in everyday life. *New Directions for Child and Adolescent Development, 1988*(41), 71-87.

Caviola, S., Mammarella, I. C., Pastore, M., & LeFevre, J.-A. (2018). Children's strategy choices on complex subtraction problems: Individual differences and developmental changes. *Frontiers in psychology, 9*(1209).

Chow, J. C., & Ekholm, E. (2019). Language domains differentially predict mathematics performance in young children. *Early Childhood Research Quarterly, 46*, 179-186.

Cobb, P., & Bauersfeld, H. (1995). *The emergence of mathematical meaning: Interaction in classroom cultures*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Colome, A., Laka, I., & Sebastián-Gallés, N. (2010). Language effects in addition: how you say it counts. *Quarterly Journal of Experimental Psychology, 63*(5), 965-983.

Cook, J. L., & Rieser, J. J. (2005). Finding the Critical Facts: Children's Visual Scan Patterns When Solving Story Problems That Contain Irrelevant Information. *Journal of educational psychology, 97*(2), 224.

Corsi, P. M. (1973). *Human memory and the medial temporal region of the brain.* ProQuest Information & Learning,

Csíkos, C., Kelemen, R., & Verschaffel, L. (2011). Fifth-grade students' approaches to and beliefs of mathematics word problem solving: a large sample Hungarian study. *ZDM, 43*(4), 561-571.

Cummins, D. D. (1991). Children's interpretations of arithmetic word problems. *Cognition and instruction, 8*(3), 261-289.

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

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*.

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. (1988). Influence of Number Size, Problem Structure and Response Mode on Children's Solutions of Multiplication Word Problems. *The Journal of Mathematical Behavior, 7*(3), 197-216.

De Corte, E. (1991). *Students' Comprehension Processes when Solving Two-Step Compare Problems*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago. https://files.eric.ed.gov/fulltext/ED331853.pdf

De Corte, E., & Verschaffel, L. (1986). *Eye-Movement Data as Access to Solution Processes of Elementary Addition and Subtraction Problems*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

De Corte, E., & Verschaffel, L. (1987). The effect of semantic structure on first graders' strategies for solving addition and subtraction word problems. *Journal for Research in Mathematics Education, 18*(5), 363-381.

De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of educational psychology, 77*(4), 460-470.

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of educational psychology, 82*(2), 359-365.

de Koning, B. B., Boonen, A. J., & van der Schoot, M. (2017). The consistency effect in word problem solving is effectively reduced through verbal instruction. *Contemporary Educational Psychology, 49*, 121-129.

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*(2), 48-55.

DeCaro, M. S., & Wieth, M. B. (2016). When higher working memory capacity hinders insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(1), 39.

Dehaene, S. (2001). Précis of the number sense. *Mind & Language, 16*(1), 16-36.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition, 1*(1), 83-120.

Dellarosa, D. (1986). A computer simulation of children's arithmetic word-problem solving. *Behavior Research Methods, Instruments, & Computers, 18*(2), 147-154.

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.

Domahs, F., Delazer, M., & Nuerk, H. C. (2006). What makes multiplication facts difficult: Problem size or neighborhood consistency? *Experimental Psychology, 53*(4), 275-282.

Domahs, F., Domahs, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Willmes, K., & Nuerk, H.-C. (2007). Neighborhood consistency in mental arithmetic: Behavioral and ERP evidence. *Behavioral and Brain Functions, 3*(1), 1-13.

Dowker, A., Bala, S., & Lloyd, D. (2008). Linguistic influences on mathematical development: How important is the transparency of the counting system? *Philosophical Psychology, 21*(4), 523-538.

Dresen, V., Pixner, S., & Moeller, K. (2020). Effects of place-value and magnitude processing on word problem solving. *Cognitive Development, 54*(2).

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologyst, 44*(2), 78-89.

Ellis, N. C. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49*(1), 234-250.

Ellis, R. (2003). *Task-based language learning and teaching*: Oxford University Press.

Evans, J. S. B. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*(4), 451-468.

Fang, Z. (2006). The language demands of science reading in middle school. *International journal of science education, 28*(5), 491-520.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*(2), 1-9.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403-434.

Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition—From single digits to arithmetic. *Quarterly Journal of Experimental Psychology, 67*(8), 1461-1483.

Fite, G. (2002). Reading and math: what is the connection? A short review of the literature. *Kansas Science Teacher, 14*, 7-11.

Foong, P. Y., & Koay, P. L. (1997). School word problems and stereotyped thinking. *Teaching and Learning, 18*(1), 73-82.

Francis, G. (1989). Thematic selection and distribution in written discourse. *Word, 40*(1-2), 201-221.

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1366.

Fuchs, L. S., Fuchs, D., Compton, D. L., Hamlett, C. L., & Wang, A. Y. (2015). Is Word-Problem Solving a Form of Text Comprehension? *Scientific Studies of Reading, 19*(3), 204-223.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., . . . Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of educational psychology, 98*(1), 29.

Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J. M., Hamlett, C. L., & Lambert, W. (2008). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of educational psychology, 100*(1), 30-47.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., . . . Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental psychology, 46*(6), 1731.

Fuchs, L. S., Geary, D. C., Fuchs, D., Compton, D. L., & Hamlett, C. L. (2016). Pathways to third-grade calculation versus word-reading competence: Are they more alike or different? *Child development, 87*(2), 558-567.

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

Fuchs, L. S., Gilbert, J. K., Powell, S. R., Cirino, P. T., Fuchs, D., Hamlett, C. L., . . . Tolar, T. D. (2016). The role of cognitive processes, foundational math skill, and calculation accuracy and fluency in word-problem solving versus prealgebraic knowledge. *Developmental psychology, 52*(12), 2085-2098.

Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlett, C. L., . . . Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to prealgebraic knowledge? *Journal of educational psychology, 106*(4), 990-1006.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., . . . Zumeta, R. O. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. *Journal of educational psychology, 101*(3), 561.

Fuentes, P. (1998). Reading comprehension in mathematics. *The Clearing House, 72*(2), 81-88.

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.

Furst, A. J., & Hitch, G. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & cognition, 28*(5), 774-782.

García, A. I., Jiménez, J. E., & Hess, S. (2006). Solving arithmetic word problems: An analysis of classification as a function of difficulty in children with and without arithmetic LD. *Journal of Learning Disabilities, 39*(3), 270-281.

Gasco, J., & Villarroel, J.-D. (2014). The motivation of secondary school students in mathematical word problem solving. *Electronic Journal of Research in Education Psychology, 12*(1), 83-106.

Gasco, J., Villarroel, J. D., & Zuazagoitia, D. (2014). Different Procedures for Solving Mathematical Word Problems in High School. *International Education Studies, 7*(7), 77-84.

Gaspard, H., Dicke, A.-L., Flunger, B., Häfner, I., Brisson, B. M., Trautwein, U., & Nagengast, B. (2016). Side effects of motivational interventions? Effects of an intervention in math classrooms on motivation in verbal domains. *AERA Open, 2*(2), 2332858416649168.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of educational psychology, 104*(1), 206.

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax, 7*(1), 55-78.

Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2014). Language affects symbolic arithmetic in children: the case of number word inversion. *Journal of experimental child psychology, 119*, 17-25.

Göbel, S. M., Shaki, S., & Fischer, M. H. (2011). The cultural number line: a review of cultural and linguistic influences on the development of number processing. *Journal of Cross-Cultural Psychology, 42*(4), 543-565.

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia, 47*(2), 604-608.

Greer, B. (1992). Multiplication and division as models of situations. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 276-295). New York: Macmillan Publishing Company.

Greer, B. (1997). Modelling reality in mathematics classrooms: The case of word problems. *Learning and Instruction, 7*(4), 293-307.

Gros, H., Sander, E., & Thibaut, J.-P. (2019). When masters of abstraction run into a concrete wall: Experts failing arithmetic word problems. *Psychonomic bulletin & review, 26*(5), 1738-1746.

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

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.

Haghverdi, M., Semnani, A. S., & Seifi, M. (2012). The relationship between different kinds of students' errors and the knowledge required to solve mathematics word problems. *Bolema: Boletim de Educação Matemática, 26*(42B), 649-666.

Halliday, M., & Matthiessen, C. (2004). An Introduction to Functional Grammar, 3rd edn, rev. *C. Matthiessen. London: Edward Arnold*.

Halliday, M., Matthiessen, C. M., & Matthiessen, C. (2014). *An introduction to functional grammar*. New York: Routledge.

Hamm, F., & van Lambalgen, M. (2002). Formal foundations for semantic theories of nominalisation. *ZAS Papers in Linguistics, 27*, 1-21.

Hardy, I., Hertel, S., Kunter, M., Klieme, E., Warwas, J., Büttner, G., & Lühken, A. (2011). Adaptive Lerngelegenheiten in der Grundschule. Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. *Zeitschrift für Pädagogik, 57*(6), 819-833.

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.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of educational psychology, 87*(1), 18-32.

Heppt, B., Haag, N., Böhme, K., & Stanat, P. (2015). The role of academic-language features for reading comprehension of language-minority students and students from low-SES families. *Reading Research Quarterly, 50*(1), 61-82.

Hines, T. M. (2013). Parity influences the difficulty of simple addition and subtraction but not multiplication problems in children. *Psychological Reports, 113*(1), 36-53.

Hinsley, D. A., Hayes, J. R., & Simon, H. A. (1977). From words to equations: Meaning and representation in algebra word problems. In P. A. Carpenter & M. A. Just (Eds.), *Cognitive Processes in Comprehension* (pp. 89-106). Hillsdale, New York: Lawrence Erlbaum Associates.

Howard, T. C. (2001). Telling their side of the story: African-American students' perceptions of culturally relevant teaching. *The Urban Review, 33*(2), 131-149.

Huber, S., Fischer, U., Moeller, K., & Nuerk, H.-C. (2013). On the interrelation of multiplication and division in secondary school children. *Frontiers in psychology, 4*(740).

Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child development, 54*(1), 84-90.

Hugerat, M. (2016). How teaching science using project-based learning strategies affects the classroom learning environment. *Learning Environments Research, 19*(3), 383-395.

Imbo, I., Vanden Bulcke, C., De Brauwer, J., & Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Frontiers in psychology, 5*(313).

Imbo, I., Vandierendonck, A., & De Rammelaere, S. (2007). The role of working memory in the carry operation of mental arithmetic: Number and value of the carry. *The Quarterly Journal of Experimental Psychology, 60*(5), 708-731.

Jaffe, K. (2006). The Importance of Mathematical Relationship Formation in Math Word Problem Solving: An Eye Movement Analysis. *Hamilton College*.

Jiménez, L., & Verschaffel, L. (2014). Development of children's solutions of non-standard arithmetic word problem solving. *Revista de Psicodidáctica, 19*(1), 93-123.

Jitendra, A. K., Petersen-Brown, S., Lein, A. E., Zaslofsky, A. F., Kunkel, A. K., Jung, P.-G., & Egan, A. M. (2015). Teaching mathematical word problem solving: The quality of evidence for strategy instruction priming the problem structure. *Journal of Learning Disabilities, 48*(1), 51-72.

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

Jordan, N. C., Kaplan, D., Nabors Oláh, L., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child development, 77*(1), 153-175.

Jordan, N. C., Levine, S. C., & Huttenlocher, J. (1995). Calculation abilities in young children with different patterns of cognitive functioning. *Journal of Learning Disabilities, 28*(1), 53-64.

Jost, K., Khader, P., Burke, M., Bien, S., & Rösler, F. (2009). Dissociating the solution processes of small, large, and zero multiplications by means of fMRI. *Neuroimage, 46*(1), 308-318.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*: Allyn & Bacon.

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental psychology, 30*(6), 949-954.

Kail, R., & Hall, L. K. (1999). Sources of developmental change in children's word-problem performance. *Journal of educational psychology, 91*(4), 660.

Kaput, J. J. (1979). Mathematics and learning: Roots of epistemological status. In J. Lochead & J. Clement (Eds.), *Cognitive Process Instruction* (pp. 289-303). Philadelphia: Franklin Institute Press.

Kaur, B., & Yeap, B. H. (2001, June). *Semantic characteristics that make arithmetic word problems difficult.* Paper presented at the In Numeracy and Beyond: Proceedings of the 24th Annual Conference of the Mathematics Education Research Group of Australasia Incorporated.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 47*(1), 145-171.

Kieran, C. (1990). Cognitive processes involved in learning school algebra. In P. N. J. Kilpatrick (Ed.), *Mathematics and Cognition* (pp. 96-112). New York: Cambridge University Press.

Kingsdorf, S., & Krawec, J. (2014). Error analysis of mathematical word problem solving across students with and without learning disabilities. *Learning Disabilities Research & Practice, 29*(2), 66-74.

Kingsdorf, S., Krawec, J., & Gritter, K. (2016). A broad look at the literature on math word problem-solving interventions for third graders. *Cogent Education, 3*(1), 1135770.

Kintsch, W. (1980). On modelling text comprehension. *22. International Congress of Psychology, Leipzig*.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological review, 92*(1), 109-129.

Klein, E., Moeller, K., Zinsberger, D., Zauner, H., Wood, G., Willmes, K., . . . Nuerk, H.-C. (2013). Object-based neglect in number processing. *Behavioral and Brain Functions, 9*(1), 5.

Klein, E., Nuerk, H.-C., Wood, G., Knops, A., & Willmes, K. (2009). The exact vs. approximate distinction in numerical cognition may not be exact, but only approximate: How different processes work together in multi-digit addition. *Brain and Cognition, 69*(2), 369-381.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. *The power of video studies in investigating teaching and learning in the classroom*, 137-160.

Kloosterman, P., & Stage, F. K. (1992). Measuring beliefs about mathematical problem solving. *School science and mathematics, 92*(3), 109-115.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & cognition, 29*(7), 1000-1009.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The journal of the learning sciences, 13*(2), 129-164.

Kyttälä, M., & Björn, P. M. (2014). The role of literacy skills in adolescents' mathematics word problem performance: Controlling for visuo-spatial ability and mathematics anxiety. *Learning and Individual Differences, 29*, 59-66.

Lachmair, M., Dudschig, C., de la Vega, I., & Kaup, B. (2014). Relating numeric cognition and language processing: Do numbers and words share a common representational platform? *Acta psychologica, 148*, 107-114.

Lave, J. (1992). Word problems: A microcosm of theories of learning. In P. Ligth & G. Butterworth (Eds.), *Context and cognition: Ways of learning and knowing* (pp. 74-92). Harvester-Wheatsheaf: Hemel Hempstead.

Lean, G. A., Clements, M. A., & Del Campo, G. (1990). Linguistic and pedagogical factors affecting children's understanding of arithmetic word problems: A comparative study. *Educational Studies in Mathematics, 21*(2), 165-191.

LeBlanc, M. D., & Weber-Russell, S. (1996). Text integration and mathematical connections: A computer model of arithmetic word problem solving. *Cognitive science, 20*(3), 357-407.

Lee, K., Ng, E. L., & Ng, S. F. (2009). The contributions of working memory and executive functioning to problem representation and solution generation in algebraic word problems. *Journal of educational psychology, 101*(2), 373-387.

Lee, K., Ng, S.-F., Ng, E.-L., & Lim, Z.-Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *Journal of experimental child psychology, 89*(2), 140-158.

Lee, K., Ng, S. F., & Bull, R. (2017). Learning and solving more complex problems: The roles of working memory, updating, and prior skills for general mathematical achievement and algebra. In *Acquisition of complex arithmetic skills and higher-order mathematics concepts* (pp. 197-220): Elsevier.

LeFevre, J.-A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 216-230.

Lepik, M. (1990). Algebraic word problems: Role of linguistic and structural variables. *Educational Studies in Mathematics, 21*(1), 83-90.

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

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction, 19*(6), 527-537.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*(2), 120-131.

Lyons, J. (1977). *Semantics* (Vol. 2). London: Cambridge University Press.

MacGregor, M., & Price, E. (1999). An exploration of aspects of language proficiency and algebra learning. *Journal for Research in Mathematics Education, 30*(4), 449-467.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Journal of educational psychology, 47*(2), 106-124.

Marsh, H. W., Martin, A. J., & Cheng, J. H. (2008). A multilevel perspective on gender in classroom motivation and climate: Potential benefits of male teachers for boys? *Journal of educational psychology, 100*(1), 78.

Martin, M. O., & Mullis, I. V. (2013). *TIMSS and PIRLS 2011: Relationships among Reading, Mathematics, and Science Achievement at the Fourth Grade--Implications for Early Learning*: ERIC.

Martin, S. A., & Bassok, M. (2005). Effects of semantic cues on mathematical modeling: Evidence from word-problem solving and equation construction tasks. *Memory & cognition, 33*(3), 471-478.

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

Mattarella-Micke, A., & Beilock, S. L. (2010). Situating math word problems: The story matters. *Psychonomic bulletin & review, 17*(1), 106-111.

Mayer, R. E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories, and templates. *Instructional Science, 10*(2), 135-175.

Mayer, R. E. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (pp. 231-273). New York: Lawrence Erlbaum Associates.

Miller, K. F., Smith, C. M., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science, 6*(1), 56-60.

Moeller, K., Klein, E., Fischer, M. H., Nuerk, H.-C., & Willmes, K. (2011). Representation of Multiplication Facts-Evidence for Partial Verbal Coding. *Behavioral and Brain Functions, 7*(1), 25.

Moeller, K., Klein, E., Kucian, K., & Willmes, K. (2014). Numerical development—from cognitive functions to neural underpinnings. *Frontiers in psychology, 5*, 1047.

Moeller, K., Klein, E., & Nuerk, H.-C. (2011a). (No) small adults: Children's processing of carry addition problems. *Developmental neuropsychology, 36*(6), 702-720.

Moeller, K., Klein, E., & Nuerk, H. C. (2011b). Three processes underlying the carry effect in addition–Evidence from eye tracking. *British Journal of Psychology, 102*(3), 623-645.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*(3), 1129-1167.

Mori, K., & Okamoto, M. (2017). The role of the updating function in solving arithmetic word problems. *Journal of educational psychology, 109*(2), 245.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2011). *The impact of reading ability on TIMSS mathematics and science achievement at the fourth grade: An analysis by item reading demands*. Paper presented at the 4th International Research Conference, Singapore. https://www.iea.nl/sites/default/files/2019-04/IRC-2013_Mullis_etal.pdf

Munez, D., Orrantia, J., & Rosales, J. (2013). The effect of external representations on compare word problems: Supporting mental model construction. *The Journal of Experimental Education, 81*(3), 337-355.

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

Muthen, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological methodology, 25*, 267-316.

Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and instruction, 9*(4), 329-389.

Neef, N. A., Nelles, D. E., Iwata, B. A., & Page, T. J. (2003). Analysis of precurrent skills in solving mathematics story problems. *Journal of Applied Behavior Analysis, 36*(1), 21-33.

Nesher, P. (1976). Three determinants of difficulty in verbal arithmetic problems. *Educational Studies in Mathematics, 7*(4), 369-388.

Nesher, P., & Teubal, E. (1975). Verbal cues as an interfering factor in verbal problem solving. *Educational Studies in Mathematics, 6*(1), 41-51.

Nichols, J. (1990). Linguistic diversity and the first settlement of the New World. *Language, 66*(3), 475-521.

Norman, S., Kemper, S., & Kynette, D. (1992). Adults' reading comprehension: Effects of syntactic complexity and working memory. *Journal of Gerontology, 47*(4), 258-265.

Nortvedt, G. A., Gustafsson, J.-E., & Lehre, A.-C. W. (2016). The importance of instructional quality for the relation between achievement in reading and mathematics. In T. N.-E. Gustafsson (Ed.), *Teacher quality, instructional quality and student outcomes* (pp. 97-113): Springer, Cham.

Nuerk, H.-C., Geppert, B. E., van Herten, M., & Willmes, K. (2002). On the impact of different number representations in the number bisection task. *Cortex, 38*(5), 691-715.

Nuerk, H.-C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2011). Extending the mental number line. *Zeitschrift für Psychologie, 219*(1), 3-22.

Nuerk, H.-C., Moeller, K., & Willmes, K. (2015). Multi-digit number processing: Overview, conceptual clarifications, and language influences. In R. C. K. A. Dowker (Ed.), *The Oxford handbook of numerical cognition*: Oxford University Press.

Nuerk, H.-C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition, 82*(1), B25-B33.

Nuerk, H.-C., Weger, U., & Willmes, K. (2005). Language effects in magnitude comparison: Small, but not irrelevant. *Brain and language, 92*(3), 262-277.

Nys, J. (2010). Complex mental arithmetic: the contribution of the number sense. *Canadian Journal of Experimental Psychology, 64*(3), 215-220.

O'Neill, D. K., Pearce, M. J., & Pick, J. L. (2004). Preschool children's narratives and performance on the Peabody Individualized Achievement Test–Revised: Evidence of a relation between early narrative and later mathematical ability. *First Language, 24*(2), 149-183.

OECD. (2010). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science (volume I)*: OECD Publishing Paris.

OECD. (2013). PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know. In: Author Washington, DC.

Oostermeijer, M., Boonen, A. J., & Jolles, J. (2014). The relation between children's constructive play activities, spatial ability, and mathematical word problem-solving performance: a mediation analysis in sixth-grade students. *Frontiers in psychology, 5*(782).

Orrantia, J., & Múñez, D. (2013). Arithmetic word problem solving: evidence for a magnitude-based mental representation. *Memory & cognition, 41*(1), 98-108.

Orrantia, J., Múñez, D., San Romualdo, S., & Verschaffel, L. (2015). Effects of numerical surface form in arithmetic word problems. *Psicológica, 36*(2), 265-281.

Orrantia, J., Rodríguez, L., Múñez, D., & Vicente, S. (2012). Inverse reference in subtraction performance: An analysis from arithmetic word problems. *The Quarterly Journal of Experimental Psychology, 65*(4), 725-738.

Orrantia, J., Rodríguez, L., & Vicente, S. (2010). Automatic activation of addition facts in arithmetic word problems. *The Quarterly Journal of Experimental Psychology, 63*(2), 310-319.

Österholm, M., & Bergqvist, E. (2012, May 11-14). *What mathematical task properties can cause an unnecessary demand of reading ability?* Paper presented at the Norma 11, The sixth Nordic Conference on Mathematics Education, , Reykjavík, Iceland.

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics, 3*(1), 285-308.

Paige, J. M. (1966). *Cognitive processes in solving algebra word problems*. New York: John Wiley & Sons.

Pape, S. J. (2003). Compare word problems: Consistency hypothesis revisited. *Contemporary Educational Psychology, 28*(3), 396-421.

Passolunghi, M. C., Cargnelutti, E., & Pellizzoni, S. (2019). The relation between cognitive and emotional factors and arithmetic problem-solving. *Educational Studies in Mathematics, 100*(3), 271-290.

Passolunghi, M. C., & Costa, H. M. (2019). Working memory and mathematical learning. In A. Fritz, V. G. Haase, & P. Räsänen (Eds.), *International Handbook of Mathematical Learning Difficulties* (pp. 407-421): Springer.

Passolunghi, M. C., & Pazzaglia, F. (2004). Individual differences in memory updating in relation to arithmetic problem solving. *Learning Individual Differences, 14*(4), 219-230.

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of experimental child psychology, 80*(1), 44-57.

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. In: American Psychological Association.

Pennock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice, 30*(3), 10-28.

Perry, B., Howard, P., & Miller, B. (1999). Some language issues in the assessment of secondary mathematics. *Making the Difference, 22*, 425-430.

Petermann, F., & Petermann, U. (2008). HAWIK-IV. *Kindheit und Entwicklung, 17*(2), 71-75.

Peterson, P. L., Fennema, E., Carpenter, T. P., & Loef, M. (1989). Teacher's pedagogical content beliefs in mathematics. *Cognition and instruction, 6*(1), 1-40.

Pixner, S., Moeller, K., Hermanova, V., Nuerk, H.-C., & Kaufmann, L. (2011). Whorf reloaded: language effects on nonverbal number processing in first grade—A trilingual study. *Journal of experimental child psychology, 108*(2), 371-382.

Pixner, S., Zuber, J., Heřmanová, V., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2011). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Research in Developmental Disabilities, 32*(6), 2683-2689.

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM, 52*(1), 33-44.

Powell, S. R., & Fuchs, L. S. (2014). Does early algebraic reasoning differ as a function of students' difficulty with calculations versus word problems? *Learning Disabilities Research & Practice, 29*(3), 106-116.

Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E., & Benholz, C. (2015). Sprachkompetenz und Mathematikleistung–Empirische Untersuchung sprachlich bedingter Hürden in den Zentralen Prüfungen 10. *Journal für Mathematik-Didaktik, 36*(1), 77-104.

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.

Rabinowitz, M., & Wooley, K. E. (1995). Much ado about nothing: The relation among computational skill, arithmetic word problem comprehension, and limited attentional resources. *Cognition and instruction, 13*(1), 51-71.

Raduan, I. H. (2010). Error analysis and the corresponding cognitive activities committed by year five primary students in solving mathematical word problems. *Procedia-Social and Behavioral Sciences, 2*(2), 3836-3838.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372.

Reusser, K. (1988). Problem solving beyond the logic of things: Contextual effects on understanding and solving word problems. *Instructional Science, 17*(4), 309-338.

Reusser, K. (1993). Tutoring systems and pedagogical theory: Representational tools for understanding, planning, and reflection in problem solving. *Computers as Cognitive Tools, 1*, 143-177.

Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. New York: Routledge.

Riley, M. S. (1984). *Development of children's problem-solving ability in arithmetic*. Retrieved from https://files.eric.ed.gov/fulltext/ED252410.pdf

Riley, M. S., Greeno, J., & Heller, J. (1983). Development of children's problem-solving ability in arithmetic. *The Development of Mathematical Thinking, 1*, 153-196.

Roettger, T. B., & Domahs, F. (2015). Grammatical number elicits SNARC and MARC effects as a function of task demands. *Quarterly Journal of Experimental Psychology, 68*(6), 1231-1248.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research, 1*(4), 201-220.

Sandilos, L. E., Rimm-Kaufman, S. E., & Cohen, J. J. (2017). Warmth and demand: The relation between students' perceptions of the classroom environment and achievement growth. *Child development, 88*(4), 1321-1337.

Sarnecka, B. W. (2014). On the relation between grammatical number and cardinal numbers in development. *Frontiers in psychology, 5*(1132).

Schäfer, R. (2015, July). *Processing and querying large web corpora with the COW14 architecture.* Paper presented at the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3), Lanchaster.

Schäfer, R., & Bildhauer, F. (2012, May). *Building large corpora from the web using a new efficient tool chain.* Paper presented at the Eighth International Conference on Language Resources and Evaluation (LREC 12).

Scheffler, T. (2005). *Nominalization in German*. Unpublished Manuscript. Department of Linguistics. University of Pennsylvania. USA. Retrieved from https://www.ling.uni-potsdam.de/~scheffler/papers/scheffler-nom.pdf

Schlager, S., Kaulvers, J., & Büchter, A. (2017, February). *Effects of linguistic variations of word problems on the achievement in high stakes tests.* Paper presented at the Tenth Congress of the European Society for Research in Mathematics Education, Dublin, Ireland.

Schley, D. R., & Fujita, K. (2014). Seeing the math in the story: on how abstraction promotes performance on mathematical word problems. *Social Psychological and Personality Science, 5*(8), 953-961.

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *LGVT 6-12: Lesegeschwindigkeits- und-verständnistest für die Klassen 6-12*: Hogrefe Göttingen.

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.

Searle, B. W., Lorton, P., & Suppes, P. (1974). Structural variables affecting CAI performance on arithmetic word problems of disadvantaged and deaf students. *Educational Studies in Mathematics, 5*(3), 371-384.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454-499.

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.

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic bulletin & review, 16*(2), 328-331.

Soltanlou, M., Pixner, S., & Nuerk, H.-C. (2015). Contribution of working memory in multiplication fact network in children may shift from verbal to visuo-spatial: a longitudinal investigation. *Frontiers in psychology, 6*.

Spanos, G., Rhodes, N., & Dale, T. (1988). Linguistic features of mathematical problem solving: Insights and applictions. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 221-240). Hillsdale, New York: Lawrence Erlbaum Associates.

Spencer, M., Fuchs, L. S., & Fuchs, D. (2020). Language-related longitudinal predictors of arithmetic word problem solving: A structural equation modeling approach. *Contemporary Educational Psychology, 60*(1), 101825.

Stacey, K. (2005). The place of problem solving in contemporary mathematics curriculum documents. *The Journal of Mathematical Behavior, 24*(3-4), 341-350.

Stacey, K. (2015). *The international assessment of mathematical literacy: PISA 2012 framework and items.* Paper presented at the Selected regular lectures from the 12th International Congress on Mathematical Education.

Staub, F., & Reusser, K. (1992, January). *The role of presentational factors in understanding and solving mathematical word problems.* Paper presented at the Meeting of the American Educational Research Association, San Francisco.

Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word problems. *Cognitive Development, 7*(2), 259-268.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta psychologica, 30*, 276-315.

Strohmaier, A. R., Lehner, M. C., Beitlich, J. T., & Reiss, K. M. (2019). Eye movements during mathematical word problem solving—global measures and individual differences. *Journal für Mathematik-Didaktik, 40*(2), 255-287.

Suviste, R., Kiuru, N., Palu, A., & Kikas, E. (2016). Classroom management practices and their associations with children's mathematics skills in two cultural groups. *Educational Psychology, 36*(2), 216-235.

Swanson, H. L. (2004). Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Memory & cognition, 32*(4), 648-661.

Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of educational psychology, 98*(2), 265.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of educational psychology, 96*(3), 471-491.

Swanson, H. L., Cooney, J. B., & Brock, S. (1993). The Influence of Working Memory and Classification Ability on Children′ s Word Problem Solution. *Journal of experimental child psychology, 55*(3), 374-395.

Swanson, H. L., & Fung, W. (2016). Working memory components and problem-solving accuracy: Are there multiple pathways? *Journal of educational psychology, 108*(8), 1153.

Swanson, H. L., Lussier, C., & Orosco, M. (2013). Effects of cognitive strategy interventions and cognitive moderators on word problem solving in children at risk for problem solving difficulties. *Learning Disabilities Research & Practice, 28*(4), 170-183.

Swanson, H. L., Lussier, C. M., & Orosco, M. J. (2015). Cognitive strategies, working memory, and growth in word problem solving in children with math difficulties. *Journal of Learning Disabilities, 48*(4), 339-358.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312.

Team, R. C. (2014). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Terao, A., Koedinger, K. R., Sohn, M.-H., Qin, Y., Anderson, J. R., & Carter, C. S. (2004). *An fMRI study of the interplay of symbolic and visuo-spatial systems in mathematical reasoning.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society, Chicago.

Thatcher, K. L., Fletcher, K., & Decker, B. (2008). Communication disorders in the school: Perspectives on academic and social success an introduction. *Psychology in the Schools, 45*(7), 579-581.

Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta psychologica, 133*(1), 90-95.

Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. In C. R. Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 158-179). Oxford: Oxford University Press.

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. *The Quarterly Journal of Experimental Psychology, 60*(1), 43-56.

Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representations in arithmetic word problem solving. *The Quarterly Journal of Experimental Psychology Section A, 58*(7), 1311-1323.

Thevenot, C., & Oakhill, J. (2006). Representations and strategies for solving dynamic and static arithmetic word problems: The role of working memory capacities. *European Journal of Cognitive Psychology, 18*(5), 756-775.

Titsworth, S., McKenna, T. P., Mazer, J. P., & Quinlan, M. M. (2013). The bright side of emotion in the classroom: Do teachers' behaviors predict students' enjoyment, hope, and pride? *Communication Education, 62*(2), 191-209.

To, V., Lê, T., & Lê, Q. (2013). A comparative study of nominalisation in IELTS writing test papers. *International Journal of Innovative Interdisciplinary Research, 1*(4), 15-21.

Tolar, T. D., Fuchs, L., Cirino, P. T., Fuchs, D., Hamlett, C. L., & Fletcher, J. M. (2012). Predicting development of mathematical word problem solving across the intermediate grades. *Journal of educational psychology, 104*(4), 1083-1093.

Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009). Jump or compensate? Strategy flexibility in the number domain up to 100. *ZDM, 41*(5), 581-590.

Träff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children. *Journal of experimental child psychology, 116*(2), 139-156.

Tyler, M. D., Tyler, L., & Burnham, D. K. (2005). The delayed trigger voice key: An improved analogue voice key for psycholinguistic research. *Behavior Research Methods, 37*(1), 139-147.

Ulu, M. (2017). Errors made by elementary fourth grade students when modelling word problems and the elimination of those errors through scaffolding. *International Electronic Journal of Elementary Education, 9*(3), 553-580.

Ulu, M., Tertemiz, N., & Peker, M. (2016). Determining the Errors of Primary School 5th Grade Students in Non-Routine Problem Solving. *Journal of Theoretical Educational Science, 9*(4), 571-605.

Van der Schoot, M., Arkema, A. H. B., Horsley, T. M., & van Lieshout, E. C. (2009). The consistency effect depends on markedness in less successful but not successful problem solvers: An eye movement study in primary school children. *Contemporary Educational Psychology, 34*(1), 58-66.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Dooren, W., Verschaffel, L., & Onghena, P. (2002). The impact of preservice teachers' content knowledge on their evaluation of students' strategies for solving arithmetic and algebra word problems. *Journal for Research in Mathematics Education*, 319-351.

Vergnaud, G. (2009). The theory of conceptual fields. *Human development, 52*(2), 83-94.

Verguts, T., & Fias, W. (2005). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Memory & cognition, 33*(1), 1-16.

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. (1990). Do nonsemantic factors also influence the solution process of addition and subtraction word problems? In E. D. C. H. Mandl, N. Bennett, & N. H. H. F. Friedrich (Ed.), *Learning and Instruction* (Vol. 2-3, pp. 415-430). Oxford: Pergamon.

Verschaffel, L., & De Corte, E. (1993). A decade of research on word problem solving in Leuven: Theoretical, methodological, and practical outcomes. *Educational Psychology Review, 5*(3), 239-256.

Verschaffel, L., De Corte, E., & Borghart, I. (1997). Pre-service teachers' conceptions and beliefs about the role of real-world knowledge in mathematical modelling of school word problems. *Learning and Instruction, 7*(4), 339-359.

Verschaffel, L., De Corte, E., Lamote, C., & Dherdt, N. (1998). The acquisition and use of an adaptive strategy for estimating numerosity. *European Journal of Psychology of Education, 13*(3), 347-370.

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., Greer, B., & De Corte, E. (2000). Making sense of word problems. *Educational Studies in Mathematics, 42*(2), 211-213.

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: a survey. *ZDM, 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.

Walkington, C., Clinton, V., & Shivraj, P. (2018). How Readability Factors Are Differentially Associated With Performance for Students of Different Backgrounds When Solving Mathematics Word Problems. *American Educational Research Journal, 55*(2), 362-414.

Weiß, R. H. (2006). *CFT 20-R: Grundintelligenz Skala 2. Revision*: Hogrefe.

Wilkins, J. L., Baroody, A. J., & Tiilikainen, S. (2001). Kindergartners' understanding of additive commutativity within the context of word problems. *Journal of experimental child psychology, 79*(1), 23-36.

Winer, B., Brown, D., & Michels, K. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

Wood, G., Nuerk, H.-C., Moeller, K., Geppert, B., Schnitker, R., Weber, J., & Willmes, K. (2008). All for one but not one for all: How multiple number representations are recruited in one numerical task. *Brain research, 1187,* 154-166.

Xin, Z., & Zhang, L. (2009). Cognitive holding power, fluid intelligence, and mathematical achievement as predictors of children's realistic problem solving. *Learning and Individual Differences, 19*(1), 124-129.

Yamamoto, S., Akao, Y., Murotsu, M., Kanbe, T., Yoshida, Y., Maeda, K., . . . Hirashima, T. (2014). *Interactive environment for learning by problem-posing of arithmetic word problems solved by one-step multiplication and division*. Paper presented at the ICCE2014, Japan.

Yarbus, A. L. (2013). *Eye movements and vision*. Boston, MA: Springer.

Zarnhofer, S., Braunstein, V., Ebner, F., Koschutnig, K., Neuper, C., Ninaus, M., . . . Ischebeck, A. (2013). Individual differences in solving arithmetic word problems. *Behavioral and Brain Functions, 9*(1), 28.

Zawaiza, T. R. W., & Gerber, M. M. (1993). Effects of explicit instruction on math word-problem solving by community college students with learning disabilities. *Learning Disability Quarterly, 16*(1), 64-79.

Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of experimental child psychology, 110*(4), 481-498.

Zhou, X., Li, M., Li, L., Zhang, Y., Cui, J., Liu, J., & Chen, C. (2018). The semantic system is involved in mathematical problem solving. *Neuroimage, 166*, 360-370.

Zimmer, K. E., Austerlitz, R., Diver, W., & Martinet, A. (1964). *Affixal Negation in English and Other Languages: An Investigation of Restricted Productivity: Supplement to Word* (Vol. 20). London: William Clowes and Sons.

Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of experimental child psychology, 102*(1), 60-77.

ACKNOWLEDGEMENT

I wish to thank all the people whose assistance was a milestone in the completion of this project.

Foremost, I would like to express my sincere gratitude to my advisor Prof. Hans-Christoph Nürk for providing guidance and feedback throughout this project. I would like to acknowledge Professor Detmar Meurers for inspiring my interest in this field. In addition, I would like to thank the LEAD Graduate School & Research Network for letting me be part of this incredible network, and for the supports through the years.

I also thank my fellow labmates, and colleagues: Dr. Mojtaba Soltanlou, Dr. Krzysztof Cipora, Dr. Christina Artemenko, Andreas Schweizer for the stimulating discussions, and for all the fun we have had in the last years. I would like to pay my special regards to my college, Dr Christina Artemenko who was instrumental in defining the path of my dissertation. For this, I am extremely grateful.

I would also like to acknowledge all individuals and the schools for their participation and engagement in the studies. Furthermore, I would like to thank the all the student assistants for their collaborative effort during data collection. Special thanks to Zoë-Lauren Kirste for proofreading my dissertation.

Last but not least, thank you my beloved daughter, Chiara, for whom I would bring the stars down from the sky.

ERKLÄRUNG

Hiermit erkläre ich, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

„ Linguistic and Arithmetic Factors in Word Problems “

selbstständig verfasst, nur die angegebenen Hilfsmittel genutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Die Arbeit ist in gleicher oder ähnlicher Form in einer anderen Prüfung noch nicht vorgelegt worden; auch wurde mit dieser Arbeit oder einer anderen Dissertation noch kein Promotionsversuch unternommen.

Tübingen, den 06.09.2020 Gabriella Daróczy