Research Article

OPEN ACCESS

Investigating Students' Processes of Noticing and Interpreting Syntactic Language Features in Word Problem Solving through Eye-Tracking

Jennifer Dröse [1]* , Susanne Prediger [2,3] , Philipp Neugebauer [2] , Renate Delucchi Danhier [2] , Barbara Mertins [2]

[1] University Paderborn, GERMANY
[2] TU Dortmund University, GERMANY
[3] IPN Leibniz Institute for Science and Mathematics Education, GERMANY
***Corresponding Author:** jennifer.droese@math.uni-paderborn.de

ARTICLE INFO	ABSTRACT
	Syntactic language features of word problems have often been identified by assessment studies and interview studies as potential obstacles in word problem solving. However, students' processes of noticing and interpreting these features have only rarely been investigated in depth because noticing is hard to investigate. Eye-tracking methodology has already been shown to be a suitable methodology for analyzing noticing other word problem features, but not yet for syntactic features. The longitudinal eye-tracking study presented here analyzes fifth graders' processes of noticing and interpreting of syntactic word problems before and after an intervention on word problem solving. Based on the eye-mind hypothesis, the study captures students' noticing through eye-tracking, whereas interpreting can be captured by their solutions. The analysis of pre-intervention data reveals that long fixations cannot be identified with successful noticing and vice versa. But when comparing the net dwell time (i.e., the time that students look at the words indicating syntactic structures) before and after the intervention, an interesting decrease is revealed. This decrease occurs for both expected and less expected syntactic structures, regardless of the accuracy of mathematization. Meanwhile, analyzing the revisits of students indicates differences between items with more or less expected structures and correctly or incorrectly mathematized items. Methodologically, the article contributes to understanding which eye-tracking measures are relevant for capturing changes in students' processes of noticing syntactic language features: The revisits and scan paths reveal more insights than the net dwell time.

INTRODUCTION

In the reading and understanding of word problems, different kinds of obstacles have been identified in earlier studies (for overviews, see Reusser, 1997; Verschaffel, Greer, & de Corte, 2000). Among these obstacles, *language obstacles* have been shown to be a crucial part, and many assessment studies contributed to identifying potential language obstacles and showing their negative impact on students' solution rates (Abedi & Lord, 2001; Dyrvold, Bergqvist, & Österholm, 2015; Haag, Heppt, Stanat, Kuhl, & Pant, 2013; further insights will be provided in Section *Syntactic Features as Potential Obstacles in Reading and Understanding Word Problems*), especially for unexpected syntactic language features, which can change the mathematization of a word problem (Boonen, van der Schoot, van Wesel, de Vries, & Jolles, 2013).

However, little is known about students' *processes of noticing and interpreting syntactic language features* (Solano-Flores, 2010). Existing qualitative video studies (e.g., Dröse, 2019; Dröse & Prediger, 2020) have been able to capture how students interpret the syntactic features, but have only achieved limited access to whether they even notice subtle differences, for instance, between an expected or less expected syntactical structure in the word problem.

In order to contribute to closing this research gap, we used eye-tracking methodology as a promising tool for capturing students' noticing processes in more detail. Eye-tracking methodology captures a person's eye movement using the reflection of infrared lights in the cornea to show where a person is looking at a particular moment in time. According to the so-called eye-mind hypothesis (Just & Carpenter, 1980), eye movements are closely related to the cognitive focus and thus provide insights into the viewer's cognitive processes. Even though the eye-mind hypothesis has also limitations (see Anderson, Bothell, & Douglass, 2004),

it has been very influential in educational research, especially in tasks related to reading. Thus, eye-tracking is a well-established method in general education research (Lai, Tsai, Yang, Hsu, Liu, Lee, Lee, Chiou, Liang, & Tsai, 2013) and is becoming increasingly popular in mathematics education research (Schindler, Lilienthal, Chadalavada, & Ögren, 2016). Some studies have investigated students' comprehension of word problems, but mainly with respect to comprehension strategies (Hegarty, Mayer, & Green, 1992; Hegarty, Mayer, & Monk, 1995; Strohmaier, Lehner, Beitlich, & Reiss, 2019; Verschaffel, de Corte, & Pauwels, 1992) rather than syntactic language obstacles.

In this paper, we present a study that explores the use of eye-tracking for investigating fifth graders' *processes of noticing syntactic language features*. The study was conducted by an interdisciplinary group of mathematics education researchers and psycholinguists working together to investigate

(1) students' eye movements as indications for their processes of noticing and interpreting syntactic features during word problem solving as reflected by their eye movements and verbal answers and

(2) changes in students' eye movement patterns after an intervention.

In the first section, the theoretical background is presented, focusing on syntactic features as potential obstacles in reading and understanding word problems and embeds the focus on the processes of noticing and interpreting into the theoretical framework of syntactic language awareness. It then outlines how the processing of syntactic features can be examined through eye-tracking. Afterwards, the methods and design of the eye-tracking intervention study are presented. Thereafter, empirical findings are provided, which are then discussed with regard to their limitations and future research.

THEORETICAL BACKGROUND

Syntactic Features as Potential Obstacles in Reading and Understanding Word Problems

For the current study, we focus on specific word problems: not rich open-modelling problems but mathematical tasks addressing real-life situations in a closed format in text form with concise problem questions—they can contain one step, two steps, or multiple steps of combining information (Dröse & Prediger, 2020). Solving arithmetic word problems requires at first reading the text to form a textbase and afterwards grasping the situation and constructing an adequate situation model, as a kind of mental model that is the interpretation done by the reader of the word problem text corresponding to the author's interpretation. The situation model is afterwards reduced to a problem model and abstracted into a mathematical model. The connective structures and solution equation derived from the mathematical models can be solved by calculation to get the numerical answer. In the end, the numerical answer needs to be interpreted to get an answer to the problem question (Cummins, Kintsch, Reusser, & Weimer, 1988; Reusser, 1989, 1993; Thevenot, 2010).

For these kinds of arithmetic word problems as well as other kinds of word problems, students' challenges in reading, understanding, and solving word problems have often been documented empirically (Daroczy, Wolska, Meurers, & Nuerk, 2015; Verschaffel et al., 2000). These obstacles can be categorized as follows (for overviews, see Daroczy et al., 2015; Reusser, 1997; Verschaffel et al., 2000):

(1) *language obstacles* in reading and understanding either lexical or syntactical features of the text

(2) *strategic obstacles* in using sustainable subject-specific reading and understanding strategies instead of strategies to process surface level characteristics of the text into a situation model

(3) *knowledge obstacles* in applying content specific general knowledge of real-world phenomena for generating a situation or problem model

(4) *conceptual obstacles* in transferring the problem model into the correct mathematical model

(5) *habitual and meta-cognitive obstacles* in having no adequate solution plan as a meta-cognitive strategy or lacking motivation for starting working on the word problem.

While Obstacles 2-5 have already been investigated in detail, this study focuses on one aspect of Obstacle 1, syntactical language obstacles. Many assessment studies and interview studies have identified language obstacles in word problems in lexical and syntactic features (Abedi & Lord, 2001; Dyrvold et al., 2015; Haag et al., 2013; Haag, Heppt, Roppelt, & Stanat, 2015; Martiniello, 2009; Prediger, Wilhelm, Büchter, Gürsoy, & Benholz, 2018; Wolf & Leon, 2009):

- *Lexical features* that may possibly be challenging in word problems*:* kind of words (academic words, unknown words), prepositions, pronouns (including agreement in number and grammatical gender), and comparative or composites

- *Syntactic features* that may possibly be challenging in word problems*:* noun phrases, relative and conditional clauses, passive voice, adverbial structures, models, and subject-object structure

Those potential obstacles might negatively influence the process of reading and understanding word problems (Haag et al., 2015; Prediger et al., 2018). The intervention focuses on different syntactic features: On the one hand, elliptic comparisons, meaning comparisons where the comparison word is left out (Schleppegrell, 2007; Halliday, 2004; Boonen et al., 2013), for instance, in the second sentence, "than each penguin chick" is left out after the comparison with "more":

- Each of the 5 penguin chicks eats 4 fishes per day. Each of the 12 grown-up penguins eats 5 fishes.

- Each of the 5 penguin chicks eats 4 fishes per day. Each of the 12 grown-up penguins eats 5 fishes more.

On the other hand, the main focus of the current paper is a typical *syntactic feature* that makes an important mathematical difference in word problems: passive voice that involves changes from subtraction to addition:

- Lina has 5 sweets. She gives 2 to her brother. How many does she have now?   → 5 - 2 = 3
- Lina has 5 sweets. She is given 2 by her brother. How many does she have now?   → 5 + 2 = 7

One syntactic feature specific to the German language is the order of subject and object (Beese & Gürsoy, 2012) (See **Figure 1**).

Less expected language structure (ULS)	Expected language structure (ELS)
Fütterung im Streichelzoo (Version A, German Original)	**Fütterung im Streichelzoo (Version B, German Original)**
Im Streichelzoo kostet eine Packung Futter für die Ziegen 2 €. Die Geschwister Paula und Jonas möchten die Ziegen füttern. Von ihren Eltern bekommt jeder 2 €. Paula nimmt zusätzlich 3 € Taschengeld mit, Jonas nur 1 €. **Ihrem** Bruder Jonas gibt **sie** 1 € ab. Wie viele Packungen Futter kann sich jeder kaufen?	Im Streichelzoo kostet eine Packung Futter für die Ziegen 2 €. Die Geschwister Paula und Jonas möchten die Ziegen füttern. Von ihren Eltern bekommt jeder 2 €. Paula nimmt zusätzlich 3 € Taschengeld mit, Jonas nur 1 €. **Ihr** Bruder Jonas gibt **ihr** 1 € ab. Wie viele Packungen Futter kann sich jeder kaufen?
Possible expression (or multi-step calculation)	**Possible expression (or multi-step calculation)**
Paula: $(2 + 3 - 1) : 2$ Jonas: $(2 + 1 + 1) : 2$	Paula: $(2 + 3 + 1) : 2$ Jonas: $(2 + 1 - 1) : 2$
Petting Zoo Problem (Version A, English adaption)	**Petting Zoo Problem (Version B, English adaption)**
One package of goat food costs € 2. Siblings Paula and Jonas want to feed the goat. Both get € 2 from their parents. Additionally, Paula brings € 3 pocket money with her, Jonas only € 1. *Her* brother Jonas, gives *she* € 1. (Literal translation, meaning "she gives him.") How many packages of goat food can each of them buy?	One package of goat food costs € 2. Siblings Paula and Jonas want to feed the goat. Both get € 2 from their parents. Additionally, Paula brings € 3 pocket money with her, Jonas only € 1. **Her** brother Jonas gives **her** € 1. How many packages of goat food can each of them buy?

Figure 1. Less expected language structure (left) and expected language structure (right) in the petting zoo problem: syntactic cases determine subject-object structure (German original from Dröse et al., 2018)

The German language allows two types of subject-object orders, the "expected structure," subject-verb-object, and the "less expected structure," object-verb-subject. The type can be identified by subtle syntactical details, namely the grammatical cases of the pronouns as shown in the examples in **Figure 1**. These subtle differences have a significant impact on the mental model of the situation and the mathematical model as they provoke using either an operation or the inverse operation (Dröse, 2019). In **Figure 1,** the change of "[To] her brother Jonas, gives she € 1" to "Her brother Jonas gives her € 1" implies the opposite exchange of money in the mental model of the situation and therefore implies either adding or subtracting € 1 for Paula or Jonas (see possible expression or multi-step calculation) in the mathematical model.

As presented above, the potentially challenging syntactic language features in word problems are well documented from the text side (see Bergqvist, Dyrvold, & Österholm, 2012; Haag et al., 2013). Apart from task-related attributes (as language features of the task) Leiss, Plath, and Schwippert (2019) also identified process attributes (as readings strategies) and personal attributes (as reading performance) as being important for constructing an adequate situation model. The students' side in particular still requires further research: Leiss, Plath, and Schwippert (2019) request developing teaching-learning arrangements for fostering understanding word problems and further research about students' cognitive processes for overcoming obstacles.

In terms of language obstacles, unfamiliar syntactic features are a kind of linguistically encoded information that is typically overlooked or not fully understood by some students (Haag et al., 2015; Gürsoy, Benholz, Renk, Prediger, & Büchter, 2013; Prediger et al., 2018), for instance, by not paying attention to linguistically encoded information: "The little details are important … because if some details are omitted, you might work with a simplified and incorrect model of the mathematical situation embedded in the word problem" (Nortvedt, 2011, p. 265).

The process of focusing one's awareness on meaning-related syntactic features and potential obstacles within this process can influence the process of reading and understanding word problems (Gürsoy et al., 2013). Boonen et al. (2013) have further shown that students' ability to deal with less expected syntactic features is a strong predictor for their word problem solving performance as a whole. This sets the focus of the current study, and with this paper we intend to contribute to reducing the research gap in examining students' processes in overcoming language obstacles by investigating two cognitive sub-processes necessary for students to deal with unexpected syntactic features. They are introduced in the next section.

Noticing and Interpreting Syntactic Features in Word Problems

Noticing and interpreting as characteristic processes in the context of language awareness

In linguistics and language education research, students' attentiveness for linguistically encoded detailed information (e.g., for less expected syntactic language features) is widely understood as a part of language awareness. The Association for Language Awareness (ALA) defines language awareness as "explicit knowledge about language, and conscious perception and sensitivity in language learning, language teaching and language use" (García, 2017, p. 264, quoted on ALA homepage). Dröse and Prediger (2020) narrowed this wide definition to syntactic awareness, in other words, the language awareness of syntactic features. The adoption of the linguistic conceptualizations has allowed the authors to draw upon linguistic results about two typical processes involved in enacting language awareness (by adapting and integrating conceptualization from Bialystok, 1986; Fandrych, 2005a,

2005b; Portmann-Tselikas, 2001): *noticing* and *interpreting.* Applying these terms to world problems is novel to our approach, and benefits of this adaptation have already been described in Dröse and Prediger (2020):

(1) *Noticing* describes the process of identifying relevant syntactic language features. In the context of this study, noticing means focusing one's attention to specific syntactic features of a word problem that are crucial for constructing a mental model of the situation. Obstacles in noticing have been documented, as students might overlook syntactic language features and not read them in detail as they do not consider them relevant (Boonen et al., 2013; Dröse, 2019; Kaulvers, Schlager, Isselbächer-Giese, & Klein, 2016).

(2) *Interpreting* describes the process of revealing the implication of a syntactic language feature for the mental model of the situation. The interpretation of those features can be based on explicit or implicit syntactic knowledge (Brimo, Apel, & Fountain, 2017). Obstacles in interpreting are traced back to missing syntactic knowledge, for instance, of second language learners (Beese & Gürsoy, 2012).

In line with Smith (2008) and Wildemann, Akbulut, and Bien-Miller (2016), we assume that general language awareness and the more specific construct of syntactic awareness can be modelled as a continuum "ranging from either no awareness at all … to a heightened, intense awareness" (Smith, 2008, p. 181). Modelling awareness in a continuum means that it is likely that awareness can be enhanced by interventions (Dröse & Prediger, 2020; Wildemann et al., 2016). The following section presents an intervention designed for this purpose that help to investigate whether changes in noticing and interpreting can be captured by eye-tracking.

Although the specific language structures in view do not exist in English, the investigation of students dealing with these structures is still relevant for an international audience as it is only exemplary for other linguistically encoded detailed information in any language.

Fostering students' processes of noticing and interpreting through the variation principle

In order to investigate how students' noticing and interpreting syntactic structures is reflected on eye-tracking data (the changeability of the eye movement pattern; see below), an intervention was developed and iteratively elaborated for fostering fifth graders' processes of noticing and interpreting (Dröse & Prediger, 2020). Alongside strategic scaffolding (see Prediger & Krägeloh, 2015; for this study the principle of strategic scaffolding is described in Dröse & Prediger, 2021), the core design principle of the intervention is the *variation principle.*

The principle of *strategic scaffolding* is used to foster reading and understanding strategies. It is based on the idea that a more skilled person supports a learner to reach the zone of proximal development, a state that the student would not be able to reach at first without help but will be able to after a sufficient period of time and support (Wood, Bruner, & Ross, 1976, p. 90). Apart from a skilled person to support a learner, in different research traditions other approaches have also been investigated, such as scaffolding through specific tools named scaffolds. Each kind of scaffold should be faded out in the end (Hannafin, Land, & Oliver, 1999).

The *variation principle* has been widely used in mathematics education, for instance, in the Chinese tradition of *Bianshi* teaching (Pang, Bao, & Ki, 2017) and the European tradition of Variation Theory (Marton & Pang, 2006). Its aim is to focus students' attention to "essential features of a concept by differentiating them from non-essential features" (Gu, Huang, & Gu, 2017, p. 14). According to the variation principle, tasks or examples are varied and compared so that students start to pay "attention to the objects of learning of a particular content systematically" (Huang, Barlow, & Prince, 2016, p. 154). This attention can either be drawn to the learning content by keeping them invariant while other features of the task vary (Pang, Bao, & Ki, 2017) or by varying features of the learning content in view while keeping other task characteristics invariant (Marton & Pang, 2006). In language education, a similar approach of contrasting language structures has been used to focus students' attention to syntactic features (Melzer, 2013).

Based on these existing approaches in mathematics and language education, the designed intervention for fifth graders (Dröse, Prediger, & Marcus, 2018) applies the variation principle to word problem text features in order to promote students' abilities to notice and interpret the slightly varied syntactic features. Following the variation principle, three main characteristics are taken into consideration for the intervention (see also Dröse & Prediger, 2021):

- Characteristic 1: *Systematically comparable word problems*: This assumes that there is a difference in relevant language features between given tasks that students can notice and identify by systematically contrasting the given word problems (Pang et al., 2017). Which language features are most relevant depends on the tasks, as we focus only on those features that need to be noticed and interpreted to form an adequate mental model of the situation.

- Characteristic 2: *Variation of language features*: In order to identify relevant language features, they are systematically varied, which means that two versions of the word problem are designed: Versions A and B. The language features in view are the only aspect varying between Versions A and B (Marton & Pang, 2006). This systematic variation combined with Characteristic 1 opens up the possibility of *noticing* the syntactic features, for instance by comparing the two versions of a word problem being presented at the same time. In **Figure 1**, students can *notice* the pronouns as being the only difference between Versions A and B.

- Characteristic 3: *Invariance of language features*: In order to generalize the impact of the language feature in view, it is kept invariant during a sequence of word problems while other features (mathematical operations, numbers, etc.) vary (Pang et al., 2017). These characteristics focus on the process of *interpreting* of syntactic features as the central theme in a reflection with students on the mental model of the situation.

An intervention using all three characteristics focuses on different syntactic features such as elliptic comparisons, which means a comparison where the comparison word is left out (Boonen et al., 2013; Halliday, 2004; Schleppegrell, 2007), for instance, in the second sentence "than each penguin chick" is left out after the comparison with "more":

- Each of the 5 penguin chicks eats 4 fishes per day. Each of the 12 grown-up penguins eats 5 fishes.
- Each of the 5 penguin chicks eats 4 fishes per day. Each of the 12 grown-up penguins eats 5 fishes more.

The intervention uses the variation principle to sensitize students for these subtle differences: By comparing the two phrases above or by comparing the two word problems from **Figure 1**, they become aware of the subtle differences in the syntactic features.

Empirical evidence for the efficacy of the designed intervention could be provided in a controlled trial with repeated measure design and paper-and-pencil pre- and post-tests for 277 fifth graders (Dröse & Prediger, 2021). The intervention group of fifth graders ($n = 118$) who followed the designed intervention revealed a high effect size for their intra-group learning gains (d = 0.87) and significantly higher learning gains than the control groups ($F_{time \times group}$ (1, 274) = 13.41, p< 0.01, η^2 = 0.02). Therefore, this intervention can be used in the current study to disentangle the students' processes of coping with the syntactic features in detail.

As the qualitative analysis of design experiments revealed (Dröse & Prediger, 2020), students' processes of *interpreting* seemed to have undergone changes while working in the intervention with varied word problems. However, the interpretation of video data revealed significant limitations in inferring students' mental processes of *noticing* from their utterances, so the next study was planned with an eye-tracking methodology, which will be presented in the next section.

Examining the Processing of Syntactic Language Features Through Eye-Tracking

Eye-tracking methodologies have been used in educational research for longer than a decade (see Lai et al., 2013, for an overview), because the method is considered to provide reliable evidence for cognitive processes related to understanding and learning. It has also been applied to focus on mathematical word problems (Hegarty et al., 1992, 1995; van der Schoot, Bakker Arkema, Horsley, & van Lieshout, 2009; Verschaffel et al., 1992). The methodologies are connected to the eye-mind hypothesis (Just & Carpenter, 1980): Individual eye fixations are assumed to be closely related to the cognitive focus of attention.

Within the last years, eye-tracking methodologies have also been increasingly applied in mathematics education research, as they allow deeper insights into the students' cognitive processes (see overview in Norqvist, Jonsson, Lithner, Qwillbard, & Holm, 2019; Strohmaier, MacKey, Obersteiner, & Reiss, 2020). Studies have been reported about proofs (Inglis & Alcock, 2012), geometry (Schindler & Lilienthal, 2019), fraction comparison strategies (Obersteiner & Tumpek, 2016), reading heuristic worked-out examples (Beitlich, Obersteiner, & Reiss, 2015), and complex word problems for adults (Strohmaier et al., 2019). In both psycholinguistics and mathematics education, reading word problems has been described as a suitable research focus for eye-tracking methodologies, as reading activities are guided by the eyes and heavily depend on the focus of attention (overview in Rayner, Pollatsek, Ashby, & Clifton, 2012). Using eye-tracking, the eye movements are captured unobtrusively while the student reads and thinks about a word problem text. Technical challenges might be that the eye tracker must be carefully calibrated beforehand and that the students cannot look away from the screen for the whole duration of the experiment.

The captured eye movements are usually not continuous and regular, thus they can be characterized by

- fixations on specific locations in a written text (which typically last about 100 to 500 ms);
- saccades, which are quick, straight movements to the next fixated location;
- revisits, which are movements going back against the reading direction to a previously read word (Rayner et al., 2012); and
- reading time, defined as the time the student needed to solve a particular task.

The temporal or counting measures usually applied for the analysis of eye movements consist of the event detection of fixations and saccades. A saccade is defined as a rapid change in the coordinates where the person is looking to, and a fixation is defined as the period between two saccades, with a minimal duration of 80 ms and a maximal dispersion of 100 pixels. Based on these parameters of automatic event detection, it is possible to obtain the following measures for the analysis of the eye-tracking data for a pre-defined area of interest (AOI):

- *Count of fixations* on the AOI
- *Count of saccades* between the AOIs
- *Count of revisits* to the AOI, in other words, each time a participant returns their gaze to an AOI after leaving it
- *Net dwell time*, as the total amount of gaze duration on an AOI, in other words, sum of all time spent looking at the AOI, continuously or discontinuously, for the whole duration of the task (Andrá, Lindström, Arzarello, Holmqvist, Robutti, & Sabena, 2013)

Previous eye-tracking studies on word problems have already shown that the *semantic structure* of a word problem can influence the eye-tracking pattern; for example, lexical structures that are inconsistent with the mathematical operation of the word problem are read more slowly, with more saccades and revisits (Hegarty et al., 1992; Verschaffel et al., 1992). In addition, unsuccessful problem solvers have been shown to adopt superficial strategies such as direct translation strategies that become visible by more total revisits to the numbers and relational terms, whereas successful problem solvers have more revisits to other relevant words in the text that might have an impact on the mental model of the situation (Hegarty et al., 1995).

With respect to language obstacles, Rayner and colleagues conducted several studies on lexical features: Inhoff and Rayner (1986) showed that the reading of words is influenced by whether the words are expectable in their context, which is also dependent on the frequency of their use. This also supposes that the less expected language structures in view might have an

impact on the reading process, as they are used less frequently. The predictability of words also has an influence on eye movement processes (Rayner & Well, 1996). Rayner and Well (1996) showed that words that are expectable in a given context are often left out or fixated on only briefly. As subject and object pronouns might be inferred from reading each of them in the order of the sentence, this might have an impact dependent on both the characteristics of the given structure (expected or less expected) and the characteristics of the reader (e.g., reading proficiency). The characteristics of the reader have also been considered when examining the noticing and interpreting of syntactic features via eye-tracking in Dröse, Prediger, Neugebauer, Delucchi Danhier, & Mertins (submitted). This study shows that the noticing is influenced by characteristics of the reader (*reading proficiency* and *multilingual background*) combined with characteristics of the syntactic structure (*expected* or *less expected*); these results about lexical features will now be investigated with respect to transferability to syntactic features.

Apart from eye-tracking studies on prototype word problems, Strohmaier et al. (2019) conducted an eye-tracking study with adults on complex word problems using global measures such as reading speed, number of saccades, proportion of regression, and mean fixation duration to compare word problems of different problem difficulty and adults with different performances. Strohmaier et al. (2019) were able to infer from their experiment that the chosen measures were sufficient for displaying the processes of encoding a text base and forming a situation or problem model. In addition, they provided empirical evidence for their hypotheses that the process of word problem solving are influenced by the difficulty of both the word problem text and individual characteristics. They were able to show that the tested adults made longer fixations in more difficult word problems. The more difficult the problems were, the more the revisits increased, but the increase was stronger for low-performing adults than for high-performing adults. This means that apart from the complexity of a word problem, the success and therefore the correctness of a mathematization and mathematical solution might also have an impact on reading processes. Therefore, it seems worthwhile to keep the performance of successful or unsuccessful problem solving as a relevant variable for investigating students' processes of noticing and interpreting. In addition, Strohmaier et al. (2019) describe their studies' results as a confirmation of Inglis and Alcock's (2012) emphasis on revisits as a relevant measure in their study on reading proofs. The current paper will therefore pay particular attention to the measure of revisits in investigating word problem reading processes via eye-tracking.

Although the above-mentioned studies revealed substantial findings on students' strategies of coping with expected or less expected semantic structures of word problems and lexical features, so far little is known about syntactic structures. From the existing results, we hypothesize that in the key sentence for the mathematization of the situation, the net dwell time on the pronouns might depend on whether they are an expected structure or not. Net dwell time was chosen since it considers length of fixations, revisits, and saccades on a target, so it is indirectly also a good measure of reading speed. Since pronouns are frequent but short words in the German language, it is easy to overlook them, unless the participants detect a less common syntactic structure. The syntactic structure of the sentences is thus systematically varied in our study.

Additionally, because all previous studies have only tested the participants once, little is known about whether the patterns of the eye movements of the participants can change over time, for instance, after an intervention, which we will call the changeability of the eye movement pattern.

Research Questions

The current state of research suggests that it might be worthwhile to also extend the eye-tracking research on word problems, which has so far mainly focused on semantic structures and lexical language features, to syntactic language awareness, namely, to students' processes of noticing syntactic features, and relate these syntactic features to the processes of interpreting language features (captured by students' utterances and mathematizations). The existing research on students' noticing of expected and less expected lexical features (Inhoff & Rayner, 1986; Rayner & Well, 1996) has suggested that this difference could also be crucial for *syntactic* features. This raises the methodological challenge of how processes of noticing syntactic features in word problems can be characterized by eye-tracking methodology and then be related to interpreting the text, as the correct interpretation cannot be derived directly from eye-tracking data. The eye-tracking study of Strohmaier et al. (2019) provided the first measures focusing on revisits. Therefore, one of the aims of the current study is to also investigate in how far the process of noticing can be captured through these eye-tracking measures.

Apart from this methodological challenge, the processes of noticing and interpreting should be investigated for word problems with expected and less expected syntactic features. Hence, the features should be compared systematically, which suggests our first research question:

RQ1.   Which differences occur in students' fixation times on pronouns in items with expected or less expected subject-object-order syntactic structures?

This comparison might reveal processes of noticing syntactic language features by taking correct and incorrect mathematizations into account during the examination of the process of interpreting syntactic language features. That means that these differences are also compared with respect to correct or incorrect problem mathematizations: We aim to capture whether students choose the correct operation, not whether they complete it. This focus resonates with the study of Strohmaier et al. (2019), which indicated the connection between the correctness of mathematization, rather than the completion of the operation, and the reading processes. From this, we derive the second research question:

RQ2.   How are these differences in expected or less expected structures connected to students' correct or incorrect mathematizations?

A second extension of the existing eye-tracking research will then be given a longitudinal perspective, asking whether an intervention can change students' eye movement patterns, and therefore possibly the processes of noticing and interpreting. This intervention has already been shown to have an impact on performance in paper-and-pencil pre- and post-tests (Dröse & Prediger,

2021) from the pre-intervention eye-tracking experiment (abbreviated pre-ET) to the post-intervention eye-tracking experiment (post-ET). This suggests the third research question:

RQ3.   How do the identified patterns change from pre-ET to post-ET?

METHODS

Figure 2 gives an overview of the research design of the longitudinal eye-tracking study with pre- and post-ET with 10 fifth graders. In order to investigate the changeability of the eye movement pattern, the longitudinal eye-tracking study with a focus sample of 10 students was embedded in a large intervention study whose entire sample consisted of 126 students. We briefly present the larger intervention, the measures for background variables, and the sampling procedure for the focus sample.
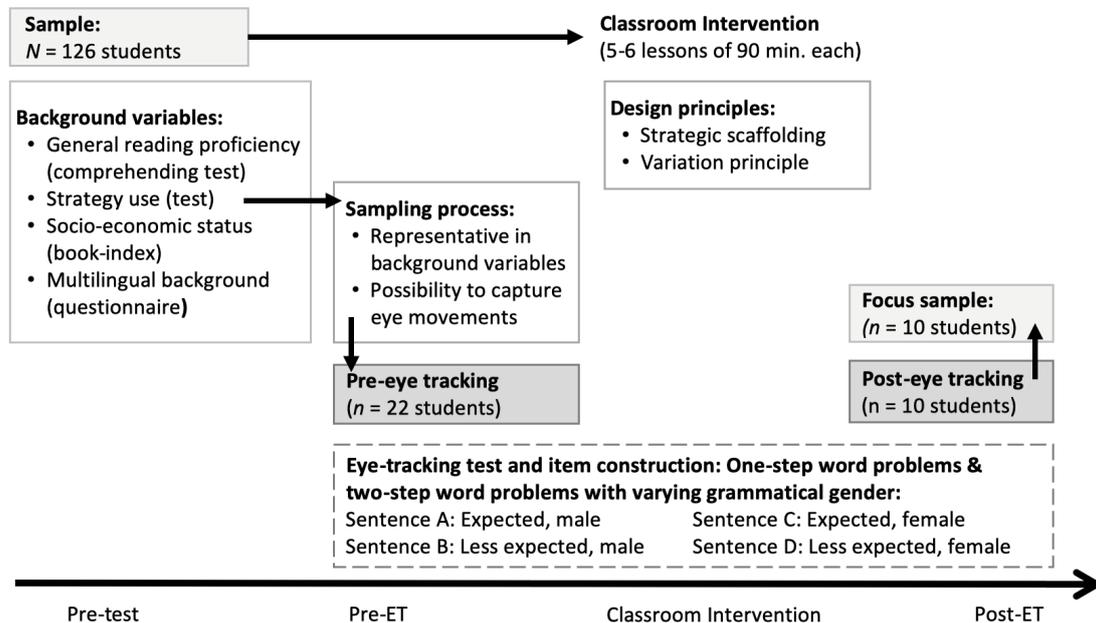

Figure 2. Design and measures of the longitudinal eye-tracking study with pre-ET and post-ET

The noticing of syntactic features was investigated using eye-tracking methodology. We present the eye-tracking tasks, the eye-tracking equipment, and the methods of data analysis.

Embedding in a Larger Intervention Study

The longitudinal eye-tracking study was embedded in a large intervention study whose entire sample consisted of 126 students. The intervention in view aimed at enhancing students' ability to deal with word problems, such as the one shown in **Figure 1**, drawing upon two design principles: strategic scaffolding (not in view of this article) and raising students' language awareness for syntactic features using the variation principle (for a detailed description of the intervention see Dröse, 2019; Dröse & Prediger, 2021).

The intervention lasted five to six classroom lessons of 90 minutes each. Significant effects on students' subject-specific comprehension of word problems have been shown using a paper-and-pencil pre- and post-test design, in which students in the intervention group showed significantly higher learning gains in solving the word problems than students in a no-treatment control group (Dröse, 2019; Dröse & Prediger, 2021).

Measures of Background Variables

The following background variables were assessed to control for differences among the students in the intervention sample and the focus sample:

- Age, gender, socio-economic status and multilingual background were captured using a self-report questionnaire. Participants' multilingual backgrounds were operationalized by listing all languages spoken at home. Socio-economic status was assessed using the book scale as an economical and reliable instrument ($r = 0.80$, Paulus, 2009).
- Reading proficiency was captured using a speed-reading comprehension test (Lenhard & Schneider, 2006), with a retest reliability of $r = 0.93$ and a maximum score of 120.
- Strategy use for reading and understanding word problems was accessed using a self-constructed test (see Dröse & Prediger, 2021). The test contains four rich word problems for which written solutions were coded and evaluated with respect to the use of the requested strategies for (S1) finding relevant information concerning the problem question, (S2) identifying the meaning of the relevant information and processing it, and (S3) the relationship between information in multistep word problems and processing it. The maximum score for these problems was 12, and the coding reached an interrater reliability of Cohen's $\kappa = 0.95$.

Participants: Intervention Sample and Focus Sample

The eye-tracking study was conducted with a focus sample of 10 fifth graders who were selected from the entire intervention sample based on the following criteria:

- Students who were representative for the whole sample concerning background variables. As the data in **Table 1** shows, the focus sample of eye-tracking students did not differ in reading proficiency, strategy use, and age from the whole intervention sample, but had a slightly higher proportion of monolingual students and students of higher socio-economic status.
- Students whose eye-movements could be captured via eye-tracking. Recordings with recording accuracy over 1.5 degrees were excluded from the study. In some cases, calibration was not successful. These cases also had to be excluded from the study.
- Students who participated in both the pre-ET and the post-ET. Due to illness, the sample of participating students was reduced. Participation was voluntary and parental consent was required.

Table 1. Background variables of the whole intervention group and the selected focus students

	Reading proficiency (max. 120) m (SD)	Strategy use (max. 12) m (SD)	Age in years m (SD)	Percentage of monolinguals / multilinguals	Socio-economic status: low / medium / high SES
Whole intervention sample (*n* = 126)	60.03 (13.51)	4.15 (2.22)	10.90 (0.72)	53%/47%	28%/25%/48%
Focus sample (*n* = 10)	59.30 (13.75)	4.85 (2.60)	10.7 (0.64)	70%/30%	0%/10%/90%

Eye-Tracking Tasks and Procedure

In the eye-tracking experiments before and after the intervention (pre-ET and post-ET), the students worked on eight word problems individually. In between the presentations of two items, fillers were shown with hidden object games. The word problems were systematically contrasted: Each item list contained (a) one-step or two-step problems (examples shown in **Figure 3**), with one sentence being varied among four possibilities, according to (b) grammatical gender of the pronouns (male/female) and (c) expected subject-object position (subject-verb-object) or less expected subject-object position (object-verb-subject). In **Figure 3** the AOI are printed in bold. The item lists have been randomly compiled out of two possibilities each for pre-ET and post-ET.

Figure 3. Test construction: Variation of eye-tracking items according to (a) one or two steps, (c) expected or less expected subject-object-order, and (c) grammatical gender of the pronouns (AOI printed in bold)

Eye-Tracking Equipment and Procedures

The eye movements were recorded with a remote eye tracker (SMI RED250) using a sampling rate of 250 Hz. We used a mobile *one-computer set-up*, comprising a laptop and a monitor to be able to record the participants at their school. The participating children sat in front of the monitor at a distance of approximately 65 cm, and their eyes were tracked without having to wear any accessories (such as eye-tracking glasses or a chin-rest), permitting them to speak freely. This was to assure that the recordings were as stress-free and non-invasive as possible since the children were quite young (10 to 11 years old). Monocular recording captured only the right eye. The monitor presented the stimuli with a ratio of 16:9 and a resolution of 19, 20 x 1080 pixel. The *one computer with remote set-up* also allowed for more accuracy and reliability in the recorded eye-tracking data than other eye-tracking equipment (e.g., eye-tracking glasses), which was of importance since the AOIs were relatively small (pronouns that are single words). The eye-tracking measures are described in the next section.

Methods of Data Analysis and Measures

The eye-tracking data was analyzed using BeGaze software (Version 3.7). The tracking ratio for the pre-ET measures was on average 90.34% (SD 8.53%) and for the post-ET 93.98% (SD 3.30%). This is an acceptable tracking ratio, especially taking into consideration that the participants were young and thus move more heavily and their eyes are closer.

In order to be able to determine fixation times and count revisits, the areas of interests (AOI) must be pre-defined (Strohmaier et al., 2019). We chose the pronouns of Sentences A, B, C, and D, as they have to be focused on to disentangle the subject-object relationship of the sentence before choosing the mathematization (addition or subtraction). The subject pronoun formed the subject AOI, the object pronoun the object AOI.

For a first rough overview of analyzing the process of noticing syntactic features, we determined the most common time measure for the fixation length because "time spent gazing in each area of interest is a relevant measure of how the participants regard and process the information" (Norqvist et al., 2019, p. 8). In line with these authors, we chose the *net dwell time* as measure best capturing the noticing of syntactic language features. The net dwell time is defined as the sum of sample durations for all gaze data samples that hit a particular AOI, in other words, the sum of the durations from all fixations and saccades that hit the AOI, in a quotient over the total viewing time for each trial. By considering relative times, we accounted for the fact that some participants took longer than others to solve the tasks (because of reading some parts of the texts again).

In the statistical analysis, the determined net dwell time was compared for all items between (1) pre-ET and post-ET, (2) subject-AOI and object-AOI, and (3) correct or incorrect mathematization. The correct of incorrect mathematization was analyzed based on the students verbally given answers. The process of interpreting has to have taken place successfully in order to mathematize the mental model correctly. Different patterns of noticing and interpreting can be identified using eye-tracking data and verbally given answers:

(1) Small amount of net dwell time combined with an incorrect mathematization: Unsuccessful noticing (without awareness of syntactic features)

(2) Small amount of net dwell time combined with a correct mathematization: Successful noticing (without awareness of syntactic features) and successful interpretation

(3) Higher amount of net dwell time combined with a correct mathematization: Successful noticing (with awareness of syntactic features) and successful interpretation

(4) Higher amount of net dwell time combined with an incorrect mathematization: Successful noticing (with awareness of syntactic features) and unsuccessful interpretation.

The comparisons of the data in order of (1) pre-ET and post-ET and (2) subject AOI and object AOI also reveal tendencies, similarities, and differences in the processing of noticing and interpreting (see Section *Results of the Eye-tracking Analyses*), even if they do not become statistically significant due to the limited number of items and participants. Because of the small sample size *t*-test results are not reported in this paper.

In a deeper analysis of the dynamics of the reading process (similar to Schindler & Lilienthal, 2019), students' *scan paths* were traced systematically from the videos. Taking into account the high relevance of revisits for reading mathematical texts (Inglis & Alcock, 2012; Strohmaier et al., 2019), scan paths allow deeper insights into students' reading processes when considering the different reading cycles.

To analyze the scan paths, it is convenient to distinguish reading cycles, in other words, each time a participant reads the text linearly from the beginning over again. In this study, students showed up to four global reading cycles of the sentence. In a second step of analysis, the revisits to the AOIs were captured for each reading cycle. This procedure not only captured revisits, but also their placement in a reading cycle, which allows for deeper comparison of pre-ET and post-ET (see Section *Results of the Eye-tracking Analyses*).

RESULTS OF THE EYE-TRACKING ANALYSES

The presented analysis starts with investigating the classical condensed measure for fixation length and average net dwell times. Afterwards, we will unpack these results in more detail, with dynamic measures counted in the students' scan paths. This illustrates the complexity of processes for three cases and then counts the revisits for all students will be presented.

Comparing Dwell Time for Different Items and Success of Mathematization

Measured net dwell times

The classical measure for capturing students' noticing is the fixation length which is operationalized here as *net dwell time* on the AOIs for the subject and object pronouns. In order to investigate the methodological question of how processes of noticing syntactic features in word problems can be characterized using eye-tracking methodology and be related to interpreting the text (since correct interpretation cannot be derived from eye-tracking data directly), we compare the net dwell time for all conditions listed in the research questions: expected or less expected subject-object order (RQ1), for item reading processes that end with students correct or incorrect mathematizations (RQ2), and for the pre-ET and post-ET (RQ3).

Table 2 presents these results for the absolute net dwell time (in milliseconds), and **Table 3** presents the same data for the relative net dwell time (i.e., in percentages of the whole reading time). Both reveal differences and similarities in the net dwell time for the subject AOI and object AOI fixation for different kinds of tasks and mathematizations.

Table 2. Absolute net dwell time on AOIs, pre-ET and post-ET, by item structure and correctness of mathematization

AOI	Pre-ET: Net dwell time on AOI (in ms)			Post-ET: Net dwell time on AOI (in ms)			Difference: pre- and post-ET net dwell time (in ms)		
	Subject	Object	Amount	Subject	Object	Amount	Subject	Object	Amount
Items with expected structure	511.61	524.03	*n = 40*	334.30	348.54	*n = 40*	-217.31	-175.49	
correct mathematizations	549.19	551.27	72.5%	329.76	365.41	75%	-219.43	-185.86	+2.5%
incorrect mathematizations	412.55	452.22	27.5%	348.03	297.92	25%	-64.55	-154.30	-2.5%
Items with less expected structure	512.04	573.21	*n = 40*	363.24	514.84	*n = 40*	-148.80	-58.37	
correct mathematization	421.69	599.26	62.5%	356.70	500., 20	77.5%	-64.99	-99.06	+15%
incorrect mathematization	662.63	529.80	37.5%	385.74	565.27	22.5%	-276.89	37.47	-15%

Table 3. Relative net dwell time on AOIs (percentage of measured whole reading time), pre-ET and post-ET, by item structure and correctness of mathematization

AOI	Pre-ET: Net dwell time on AOI (in %)			Post-ET: Net dwell time on AOI (in %)			Difference: pre-post-ET Net dwell time (in %)		
	Subject	Object	Amount	Subject	Object	Amount	Subject	Object	Amount
Items with expected structure	1.8%	1.8%	*n = 40*	1.8%	1.9%	*n = 40*	+ 0.0%	+ 0.1%	
correct mathematizations	1.9%	1.9%	72.5%	1.9%	2.0%	75%	+ 0.0%	+ 0.1%	+1%
incorrect mathematizations	1.3%	1.4%	27.5%	1.8%	1.5%	25%	+ 0.5%	+ 0.1%	-1%
Items with less expected structure	2.0%	2.2%	*n = 40*	1.8%	2.5%	*n = 40*	- 0.2%	+ 0.3%	
correct mathematization	1.7%	2.4%	62.5%	1.8%	2.5%	77.5%	+ 0.1%	+ 0.1%	+15.5%
incorrect mathematization	2.5%	2.0%	37.5%	1.7%	2.6%	22.5%	- 0.8%	+ 0.6%	-15.5%

A first look shows that the absolute net dwell time for an AOI varies between 412 ms and 662 ms, which corresponds to a relative net dwell time of between 1.3% and 2.5%. This means that the students fixated on the subject or object pronouns in 1.3% to 2.5% of their measured whole reading time.

Analysis of differences and similarities in net dwell time on AOI in pre-ET

A comparison of the items with expected or less expected subject-object order in the pre-ET shows differing success in mathematization, with 72.5% correct mathematizations for items with expected structure and 62.5% for items with less expected structure. With respect to the relative net dwell time, the average measures seem to be similar: students fixated on the AOI in items with expected structures slightly shorter (1.8% for subject and object AOIs) than in items with less expected structures (2.0% on subject AOIs and 2.2% on object AOIs).

However, the differences get much larger when processes of reading items with correct mathematizations are distinguished from those with incorrect mathematizations: For items with expected structures that were correctly mathematized, students fixated on both AOIs equally long (1.9%), but on the object AOIs much longer in items with less expected structures (1.7% on subject and 2.4% on object). In processes of reading items with expected structures that were incorrectly mathematized, students fixated on both AOIs for equally short amounts of time (1.3% subject and 1.4% on object), but on subject AOIs much longer in items with less expected structures (2.5% subject and 2.0% object). This means that in correctly mathematized items with less expected structures, the net dwell time on object AOIs was higher than for subject AOIs, whereas for incorrectly mathematized items the relative net dwell time on subject AOIs was higher than on object AOIs. Therefore, small differences across different items can be identified with regard to the success of mathematization. Hence, we see differentiated patterns that do not allow identification of long fixation with successful noticing or vice versa. Rather, the success of mathematization seems to depend on what to fixate on and in which context.

Analysis of changes in AOI fixation from pre-ET to post-ET

When comparing pre-ET to post-ET, the percentage of correct mathematizations increases only slightly for items with expected structures (from 72.5% to 75%), but much more for items with less expected structures (from 62.5% to 77.5%), which are even mathematized slightly better than the items with expected structures.

Additionally, **Table 2** shows the dominant pattern of decreasing absolute net dwell time for all kinds of items (with a decrease varying from 58.37 ms to 276.89 ms). However, these decreases correspond to an overall decrease in the measured whole reading time, so they are not reflected by the relative net dwell times in **Table 3**. The relative net dwell times vary from -0.8% to + 0.6%, and students seem to make the longest fixations on the object AOIs for items with less expected structures (2.5 and 2.6% relative net dwell time for correctly and incorrectly mathematized items, respectively).

For incorrectly mathematized items, the relative net dwell time on the AOIs is higher in post-ET than in pre-ET. This might indicate that students pay more attention to the syntactic features in view; therefore, it is likely that processes of noticing take place. However, the students do not always interpret the features correctly, so the items are not always mathematized correctly. Hence, these results suggest that students can notice syntactic features after the intervention, but more than 25% of the items are still not interpreted.

This might indicate that students become faster in noticing syntactic features and that many students become more successful in interpreting and mathematizing syntactic features. The point of increased speed in particular has not yet been able to be shown in earlier publications (Dröse & Prediger, 2020).

As this analysis did not distinguish the students but rather grouped the processes of items according to the correctness of the mathematization, the next section changes the perspective and considers students.

Case Analysis of Three Students' Noticing and Interpreting Syntactic Features

For a deeper and student-related answer to the methodological question of *how processes of noticing syntactic features in word problems can be characterized by eye-tracking methodology and be related to interpreting the text (since correct interpretations cannot be derived from eye-tracking data directly*, we start by presenting three cases of students. The three focus students, Lars, Timo, and Mike, were chosen because they reveal different but typical processes of noticing and interpreting in their AOI fixation in the second, third, and fourth readings of the word problem Sentences A, B, C, and D.

In order to examine the presented average values in terms of the dynamics of the reading process, **Figure 4** shows different scan paths during students' reading of items with less expected structures (Sentence Types B and D) in pre-ET and post-ET. In the scan paths, fixations are represented by circles, the bigger the circle the longer the fixation. Saccades between fixations are represented by straight lines (even though the saccade itself was not necessarily linear). The main purpose of this representation is to provide a visual representation of the number, order, and length of fixations in an intuitive way.
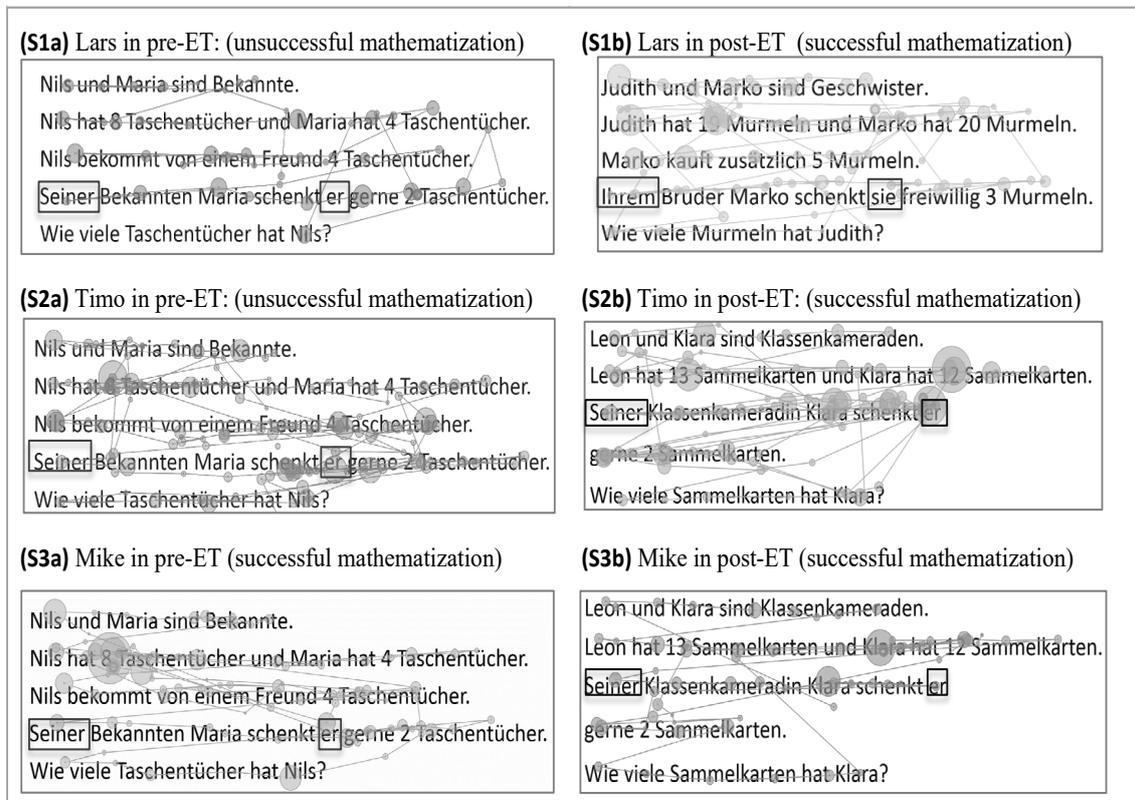

Figure 4. Six scan paths while reading items with less expected structures in pre- and post-ET

The presented scan paths illustrate typical reading processes of items with less expected structures, which can be summarized and contrasted with respect to the visible processes of noticing and interpreting (Dröse & Prediger, 2020) as follows:

In the pre-ET, three different patterns can be seen taking the scan paths and the mathematizations of the pronouns into considerations:

- (S1a): Scan path S1a shows that Lars did not focus on the AOIs "Seiner" and "er" in the second read and afterwards, and he also did not mathematize their meaning correctly. We infer that he did not *notice* the pronouns as relevant for the mathematization of the whole word problem. As a consequence, he could not *interpret* them correctly. [Pattern 1 of noticing and interpreting]

- (S2a): Timo focused on both AOIs during his second read and afterwards, but he did also not mathematize their meaning correctly. Possibly, in contrast to Lars, he did *notice the pronouns*; however, he might have had problems in *interpreting* the meaning of their grammatical cases as indicating subject and object. [Pattern 4 of noticing and interpreting]

- (S3a): Mike focused on the AOI "er" during his second read. He showed a scan path similar to that of Timo, but he also mathematized the sentence correctly. This implies that he was able to *notice* the pronouns and *interpret* their grammatical cases correctly. [Pattern 3 of noticing and interpreting]

In order to get first insights into RQ3 (*How do the identified patterns change from the pre-ET to the post-ET?*), the three pre-ET scan paths can be compared to the post-ET scan paths. In post-ET, all three students focus on the AOIs during second read, whereas in the third and fourth read, they rarely fixate on the AOIs. All three students solve between one and all items with less expected structures correctly. In those reading processes the scan paths reveal the following patterns:

- (S1b): In post-ET, Lars focused on the AOIs and mathematized them correctly. This indicates that he was now able to *notice* and *interpret* them during the process of word problem solving. [Pattern 3 of noticing and interpreting]

- (S2b): Timo incorrectly mathematized most of the items with less expected structures in post-ET. But for those items that he mathematized correctly, his focus on the pronouns was visible during the second, third, and fourth reads. This indicates that he *noticed* the pronouns, but required high effort to *interpret* those structures correctly. [Patterns 3 and 4 of noticing and interpreting]

- (S3b): Mike focused on the AOIs, but only in the first read, and mathematizes them correctly. This shows that he *noticed* and *interpreted* the pronouns during the first read and even did not need to refocus on them in the further reads. [Pattern 2 of noticing and interpreting]

So far, we have only investigated six scan paths from **Figure 4**, two for each of the three students. **Table 4** provides two generalizations: (a) it gives an overview of the three students' revisits of *all* 48 items (eight in pre-ET and eight in each post-ET), and (b) expands these results to the other seven students. The table shows the percentages of items with revisits to subject and/or object AOIs during the second, third, and fourth reads and the percentage of correct mathematizations. **Table 4** also allows comparisons between pre-ET and post-ET and between items with more and less expected structures and provides the data on the percentage of items in which the interpretation of the syntactic features was mathematized correctly. The order of students is arranged based on their percentage of correctly mathematized word problems with less expected structures. The colors show nearly no, slight, and high numbers of revisits to the AOIs.

Table 4. Revisits to areas of interest (AOIs: pronouns for subject and objects) during word problem reading: Overview of all items of all 10 students, showing similar patterns to the focus students, Lars, Timo, and Mike (in bold type)

		Pre-ET				Post-ET			
		Items with revisits to AOI during:			Correct mathematization	Items with revisits to AOI during:			Correct mathematization
		2nd read	3rd read	4th read		2nd read	3rd read	4th read	
Lars	**expected**	0%	0%	0%	25%	25%	0%	0%	100%
	less expected	25%	0%	0%	25%	25%	0%	0%	75%
Eric	expected	50%	25%	0%	25%	25%	0%	0%	50%
	less expected	25%	0%	0%	25%	25%	0%	0%	50%
Lea	Expected	25%	25%	25%	75%	0%	0%	0%	75%
	less expected	25%	25%	0%	25%	25%	0%	0%	50%
Timo	**expected**	75%	50%	25%	75%	50%	0%	0%	100%
	less expected	50%	50%	25%	25%	50%	25%	0%	25%
Fynn	expected	75%	0%	0%	50%	25%	0%	0%	25%
	less expected	50%	25%	25%	50%	0%	0%	0%	75%
Alex	expected	50%	0%	0%	50%	0%	0%	0%	100%
	less expected	75%	25%	0%	50%	0%	0%	0%	100%
Ahmed	expected	25%	25%	0%	75%	25%	0%	0%	75%
	less expected	25%	25%	0%	75%	0%	0%	0%	100%
Sina	expected	50%	0%	0%	25%	25%	0%	0%	100%
	less expected	0%	0%	0%	75%	0%	0%	0%	75%
Simon	expected	0%	0%	0%	50%	50%	0%	0%	100%
	less expected	25%	0%	0%	100%	25%	0%	0%	100%
Mike	**expected**	50%	25%	25%	75%	25%	0%	0%	75%
	less expected	75%	0%	0%	100%	25%	25%	0%	100%

The cases of the three students illustrate how processes of noticing and interpreting can be revealed through eye-tracking in pre-ET. Similarities and differences can be found among all 10 students:

(1) The counted revisits in Lars' pre-ET was low, which means he refocused the subject/object pronouns during the process of reading and solving the word problems. This indicates that he did not pay particular attention to the syntactic features in view. In connection with the low rate of correct mathematizations, this indicates (as described above) that those features were not noticed in detail. For Lars, this equally applies for items with more and less expected structures [Pattern 1 of noticing and interpreting]. Sina and Simon show the same pattern of revisits to AOIs in their scan paths, but in contrast to Lars, they correctly mathematized more than 75% of the word problems with less expected structures. This indicates that they were able to interpret those structures correctly and therefore also must have noticed them. However, the structures might be so familiar to them that they did not need to fixate again to notice them [Pattern 2 of noticing and interpreting].

(2) Timo's revisits in the pre-ET show a different pattern in his re-reading processes: In items with more and less expected structures, Timo shows many revisits to the chosen AOIs. This suggests that Timo already paid particular attention to those features. Only for items with expected structures did he often mathematize correctly [Pattern 3], whereas the items with less expected structures seem to have been noticed but not correctly interpreted (as already described for the exemplary scan path) [Pattern 4]. We infer that *noticing* was likely to take place successfully, whereas *interpreting* was not successful for items with less expected structures. Eric, Lea, Fynn, and Alex show a similar pattern of revisiting the AOIs, but mathematized only up to 50% of the word problems with less expected structures correctly.

(3) Mike's revisits and percentage of correct mathematizations are different from the other ones. He shows many revisits for items with expected structures and less expected structures, even if not as many as Timo. He also mathematized nearly all items correctly. Hence, the indications from the exemplary scan path that the process of *noticing* and the process of

interpreting have been gone through successfully are likely to be extendable to all items. Ahmed shows a similar pattern, his number of revisits is also not as high as Mike's during the second read [Pattern 3].

This comparison shows that the revisits cannot be interpreted alone but in connection with the correct mathematizations. Taking both into consideration, patterns different from the original Patterns 1-4 might be revealed.

Extending the analysis of revisits from the pre-ET to the post-ET, similar changes become visible for most of the 10 students:

- At first, nearly all revisit the subject and object pronouns in the second read of the post-ET (apart from Alex), which indicates that they pay attention to those syntactic features and therefore possibly notice them (as described in Lars' scan path above). Timo, Mike, and Simon show a higher percentage of revisits than the other seven students.
- Second, all of them increased the percentage of correct mathematizations for more and less expected structures, which is considered an indicator that the students' processes of *noticing* and *interpreting* were completed more successfully in post-ET than in pre-ET. These findings resonate with the qualitative analysis of students' utterances in design experiments in Dröse and Prediger (2020), and Dröse (2019).

Thus, capturing the revisits in eye-tracking scan paths seems to provide valid information and seems to enrich and deepen the insights into students' processes of noticing and interpreting without communicative or cooperative situations and even without self-report by students.

Apart from the individual changes in noticing and interpreting syntactic features, the presented case studies hint at possible existing differences between items with more and less expected structures and between correctly and incorrectly mathematized items. These distinctions will be further examined in the next section for the overall development in noticing and interpreting via eye-tracking.

CONCLUSION AND DISCUSSION

Results of the Eye-Tracking Study

Methodological challenge: How can processes of noticing syntactic features in word problems be characterized by eye-tracking methodology and be related to interpreting the text?

The results indicate that the processes of noticing and interpreting of syntactic features can only be partially captured by net dwell time; it is better characterized by revisits and scan paths: Revisits have been shown to be an important measure for investigating the encoding towards a textbase and the construction of a situation or problem model by Inglis and Alcock (2012) and Strohmaier et al. (2019) and, on the other hand, by the analysis of the students' scan paths. Both measures (revisits and analysis of scan paths) seem more adequate than the measure of average net dwell time. Whereas the average net dwell time revealed no big changes when measured relative to the whole reading time, analyzing the scan paths and the revisits in second, third, and fourth reads of the word problem text revealed interesting patterns. This methodological lesson learned corresponds to the methodological conclusions by Inglis and Alcock (2012) on reading proofs and Strohmaier et al. (2019) on adults' comprehension of word problems.

From the perspective of syntactic language awareness, the eye-tracking data can be used to measure students' noticing. The process of noticing can be revealed though revisits to pronouns, combined with the correct or incorrect mathematization that indicates the processes of interpreting. The percentages of revisits in pre-ET show that there are students who do not notice the relevant syntactic features as they do not revisit them and do not mathematize them correctly (e.g., Lars). These results could strengthen earlier hypotheses gained in purely observational studies that syntactic language features might be overlooked by students (Gürsoy et al., 2013; Haag et al., 2015).

Furthermore, there are students (e.g., Timo) who revisit the pronouns and therefore pay special attention to the relevant syntactic structures, but nonetheless do not necessarily mathematize them correctly. This indicates that noticing a feature can be clearly separated from interpreting it successfully. In this aspect also, the eye-tracking data resonate with hypotheses gained in observational studies (Dröse & Prediger, 2020).

Lastly, there are students like Mike who are already capable of noticing and interpreting the features correctly in pre-ET. In post-ET, nearly all of the 10 students focused on the pronouns during the second read, which indicates that after the intervention, they paid more attention to the linguistically encoded detailed information and therefore noticed it more often. In addition, there were more correct mathematizations of the structures, indicating that the processes of interpreting were being completed successfully. This deepens the insights into the evaluation of the intervention as already presented in Dröse and Prediger (2021). However, there seem to be differences in items with more and less expected structures and in items with correct and incorrect mathematizations. The positive effects of the intervention have been traced back to previous studies in the field of bianshi teaching and variation theory (Gu, Huang, & Gu, 2017; Huang, Barlow, & Prince, 2016; Marton & Pang, 2006; Pang, Bao, & Ki, 2017).

RQ1. + RQ2. Which differences occur in students' fixation of pronouns in items with more or less expected structures? And how are these differences in more or less expected structures connected to students' correct or incorrect mathematizations?

Items with expected syntactic structures are focused on for slightly shorter amounts of time than items with less expected structures in pre-ET. This resonates with previous findings concerning the influence of the expectedness and predictability of lexical features on reading processes and eye movements (Rayner & Well, 1996). For items with expected structures, both AOIs are focused on for equally lengths of time regardless of correct or incorrect mathematization. This indicates that both AOIs are equally considered and that the success of mathematizations might be mostly attributed to processes of interpreting the syntactic

features. For items with less expected structures, correctly mathematized items have a slightly higher focus on the object AOI, whereas incorrectly mathematized items have a slightly higher net dwell time on the subject AOI. These findings extend the findings of Strohmaier et al. (2019) concerning the connections between the performance in word problem solving and eye movements. To interpret those findings, the analysis had to be deepened to the revisits in the scan paths.

RQ3. How do the identified patterns change from the pre-ET to the post-ET?

Comparing pre-ET and post-ET reveals that the reading process of the given word problems got faster, as there was a reduction of net dwell time on all AOIs. Meanwhile, the relative dwell time on the AOIs did not decrease, but there was a slight tendency for an increasing relative dwell time on the pronouns. This might indicate that more processes of noticing took place. In addition, more correct mathematizations can be documented, revealing more processes of successful interpretation of the syntactic structures. Nevertheless, the students still did not all interpret the syntactic features correctly. These results underpin and extend the findings from previous studies (Dröse, 2019; Dröse & Prediger, 2021).

Limitations and Future Research

The presented study is limited due to the small sample size and the limited number of eight items. Future research will have to extend the eye-tracking study to a higher number of participants in pre-ET and post-ET. Furthermore, syntactic features other than subject-object positions (e.g., active-passive construction and comparatives) should be taken into account. In the designed teaching-learning arrangement, elliptic omission of comparative phrases is also chosen as a learning content. The video-based qualitative analysis by Dröse (2019) indicates that there are similarities but also differences in the students' learning process for both syntactic structures. Therefore, these differences might be further investigated using eye-tracking. Extending the study to other syntactic features also implies extending it to other grades and age levels and other mathematical learning content (see Prediger & Zindel, 2017, for functional relationships—without eye-tracking).

REFERENCES

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219-234. https://doi.org/10.1207/S15324818AME1403_2

Anderson, J., Bothell, D., & Douglass, S. (2004). Eye Movements Do Not Reflect Retrieval Processes Limits of the Eye-Mind Hypothesis. *Psychological Science, 15*(4), 225-231. https://doi.org/10.1111/j.0956-7976.2004.00656.x

Andrá, C., Lindström, P., Arzarello, F., Holmqvist, K., Robutti, O., & Sabena, C. (2013). Reading mathematics representations: an eye-tracking study. *International Journal of Science and Mathematics Education, 13*(2), 237-259. https://doi.org/10.1007/s10763-013-9484-y

Beese, M., & Gürsoy, E. (2012). Bezüge im Deutschen und im Türkischen herstellen. Sprachliche Stolpersteine beim Mathematiklernen für zweisprachige Lernende. *Praxis der Mathematik in der Schule, 54*(45), 34-37.

Beitlich, J., Obersteiner, A., & Reiss, K. (2015). How do secondary school students make use of different repre-sentation formats in heuristic worked examples? - An analysis of eye movements. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of PME39* (Vol. 2, pp. 97-104). Hobart: PME.

Bergqvist, E., Dyrvold, A., & Österholm, M. (2012). Relating Vocabulary in Mathematical Tasks to Aspects of Reading and Solving. In C. Bergsten, E. Jablonka, & M. Raman (Eds.), *Evaluation and comparison of mathe-matical achievement: Dimensions and perspectives: Proceedings of MADIF 8, The Eighth Mathematics Education Research Seminar, Umea*, January, 24-25, 2012 (pp. 61-70). SMDF.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development, 57*(2), 498-510. https://doi.org/10.2307/1130604

Boonen, A. J. H., 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. https://doi.org/10.1016/j.cedpsych.2013.05.001

Brimo, D., Apel, K., & Fountain, T. (2017). Examining the contributions of syntactic awareness and syntactic knowledge to reading comprehension. *Journal of Research in Reading, 40*(1), 57-74. https://doi.org/10.1111/1467-9817.12050

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20*(4), 405-438. https://doi.org/10.1016/0010-0285(88)90011-4

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: a review of linguistic and nu-merical factors contributing to their difficulty. *Frontiers in Psychology, 6*(348), 1-13. https://doi.org/10.3389/fpsyg.2015.00348

Dröse, J. (2019). *Textaufgaben lesen und verstehen lernen. Entwicklungsforschungsstudie zur mathematikspezifischen Leseverständnisförderung* [Reading and comprehending word problems: Design research study on mathematics-specific reading comprehension]. Heidelberg: Springer Spektrum. https://doi.org/10.1007/978-3-658-27850-2

Dröse, J., & Prediger, S. (2020). Enhancing fifth graders' awareness of syntactic features in mathematical word problems: A Design Research study on the variation principle. *Journal für Mathematik-Didaktik, 41*(2), 391-422. https://doi.org/10.1007/s13138-019-00153-z

Dröse, J., & Prediger, S. (2021). Identifying obstacles is not enough for everybody – Differential efficacy of an intervention fostering fifth graders' comprehension for word problems. *Studies in Educational Evaluation, 68*(100953), 1–15. https://doi.org/10.1016/j.stueduc.2020.100953

Dröse, J., Prediger, S., & Marcus, A. (2018). *Strategien zum Verstehen von Textaufgaben – Fach- und sprachintegriertes Unterrichtsmaterial in Basis- und Regelfassung*. Retrieved from Open educational resources on www.sima.dzlm.de/um/5-001

Dröse, J., Prediger, S., Neugebauer, P., Delucchi Danhier, R., & Mertins, M. (submitted). Eye-Tracking-Studie zu Textaufgaben in Klasse 5: Syntaktische Strukturen bemerken und interpretieren bei verschiedenen Sprachprofilen. Submitted Manuscript.

Dyrvold, A., Bergqvist, E., & Österholm, M. (2015). Uncommon vocabulary in mathematical tasks in relation to demand of reading ability and solution frequency. *Nordic Studies in Mathematics Education, 20*(1), 5-31.

Fandrych, C. (2005a). Ordnung und Variation in Satz und Text. Wortstellung entdecken, erkunden, erproben. *Fremdsprache Deutsch, 32*, 5-11. https://doi.org/10.37307/j.2194-1823.2005.32.03

Fandrych, C. (2005b). Schauplatz: Wortstellung. Wichtige Regeln und Tendenzen. *Fremdsprache Deutsch, 32*, 12-18. https://doi.org/10.37307/j.2194-1823.2005.32.04

García, O. (2017). Critical multilingual language awareness and teacher education. In J. Cenoz, S. May & D. Gorter (Eds.), *Language Awareness and Multilingualism* (pp. 263-280). Cham: Springer. https://doi.org/10.1007/978-3-319-02240-6_30

Gu, F., Huang, R., & Gu, L. (2017). Theory and development of teaching through variation in mathematics in China. In R. Huang & Y. Li (Eds.), *Teaching and learning mathematics through variation. Confucian heritage meets western theories* (pp. 13-41). Rotterdam: Sense Publishers. https://doi.org/10.1007/978-94-6300-782-5_2

Gürsoy, E., Benholz, C., Renk, N., Prediger, S., & Büchter, A. (2013). Erlös_ Erlösung? - Sprachliche und konzeptuelle Hürden in Prüfungsaufgaben zur Mathematik [Redemption_ redemption? - Linguistic and conceptual hurdles in mathematics exam questions]. *Deutsch als Zweitsprache, 13*(1), 14-24.

Haag, N., Heppt, B., Roppelt, A., & Stanat, P. (2015). Linguistic simplification of mathematics items: Effects for language minority students in Germany. *European Journal of Psychology of Education, 30*(2), 145-167. https://doi.org/10.1007/s10212-014-0233-6

Haag, N., Heppt, B., Stanat, P., Kuhl, P., & Pant, H. A. (2013). Second language learners' performance in mathe-matics: Disentangling the effects of academic language features. *Learning and Instruction, 28*, 24-34. https://doi.org/10.1016/j.learninstruc.2013.04.001

Halliday, M. A. K. (2004). *An introduction to functional grammar*. New York: Oxford University Press. https://doi.org/10.4324/9780203783771

Hannafin, M., Land, S., & Oliver, K. (1999). Open Learning Environments: Foundations, Methods, and Models. In C. M. Reigeluth (Ed.), *Instructional-design Theories and Models* (pp. 115-140). Mahwah: Lawrence Erlbaum.

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of Arithmetic Word Problems: Evidence From Students' Eye Fixations. *Journal of Educational Psychology, 84*(1), 76-84. https://doi.org/10.1037/0022-0663.84.1.76

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. https://doi.org/10.1037/0022-0663.87.1.18

Huang, R., Barlow, A. T., & Prince, K. (2016). The same tasks, different learning opportunities: An analysis of two exemplary lessons in China and the U.S. from a perspective of variation. *The Journal of Mathematical Behavior, 41*, 141-158. https://doi.org/10.1016/j.jmathb.2015.12.001

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Re-search in Mathematics Education, 43*(4), 358-390. https://doi.org/10.5951/jresematheduc.43.4.0358

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*, 431-440. https://doi.org/10.3758/BF03208203

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354. https://doi.org/10.1037/0033-295X.87.4.329

Kaulvers, J., Schlager, S., Isselbächer-Giese, A., & Klein, M. (2016). Entwickeln, Beraten, Unterstützen - QUA-LIS. Sprachliche Hürden in Mathematikaufgaben. *Schule NRW, 07/08*, 16-19.

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10*, 90-115. https://doi.org/10.1016/j.edurev.2013.10.001

Leiss, D., Plath, J., & Schwippert, K. (2019). Language and Mathematics - Key factors influencing the compre-hension process in reality-based tasks. *Mathematical Thinking and Learning, 21*(2), 131-153. https://doi.org/10.1080/10986065.2019.1570835

Lenhard, W., & Schneider, W. (2006). ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler. In M. Hasselhorn, H. Marx, & W. Schneider (Eds.), *Deutsch Schultest*. Göttingen: Hogrefe.

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment, 14*(3-4), 160-179. https://doi.org/10.1080/10627190903422906

Marton, F., & Pang, M.F. (2006). On some necessary conditions of learning. *The Journal of the Learning Sciences, 15*(2), 193-2, 20. https://doi.org/10.1207/s15327809jls1502_2

Melzer, F. (2013). Modellierung, Diagnose und Förderung von Sprachbewusstheit in der Sekundarstufe. In S. Gailberger & F. Wietzke (Eds.), *Handbuch kompetenzorientierter Deutschunterricht* (pp. 300-321). Weinheim: Beltz.

Norqvist, M., Jonsson, B., Lithner, J., Qwillbard, T., & Holm, L. (2019). Investigating algorithmic and creative reasoning strategies by eye tracking. *The Journal of Mathematical Behavior, 55*, 1-14. https://doi.org/10.1016/j.jmathb.2019.03.008

Nortvedt, G. A. (2011). Coping strategies applied to comprehend multistep arithmetic word problems by stu-dents with above-average numeracy skills and below-average reading skills. *The Journal of Mathematical Behavior, 30*(3), 255-269. https://doi.org/10.1016/j.jmathb.2011.04.003

Obersteiner A., & Tumpek, C. (2016). Measuring fraction comparison strategies with eye-tracking. *ZDM Mathematics Education, 48*, 255-266. https://doi.org/10.1007/s11858-015-0742-z

Pang, M. F., Bao, J., & Ki, W. W. (2017). "Bianshi" and the variation theory of learning: Illustrating two frameworks of variation and invariance in the teaching of mathematics. In R. Huang & Y. Li (Eds.), *Teaching and learning mathematics through variation. Confucian heritage meets western theories* (pp. 43-67). Rotterdam: Sense. https://doi.org/10.1007/978-94-6300-782-5_3

Paulus, C. (2009). Die „Bücheraufgabe" zur Bestimmung des kulturellen Kapitals bei Grundschülern. Retrieved on 20 June 2019 from http://hdl.handle.net/20.500.11780/3344

Portmann-Tselikas, P. R. (2001). Sprachaufmerksamkeit und Grammatiklernen. In P. R. Portmann-Tselikas & S. Schmölzer-Eibinger (Eds.), *Grammatik und Sprachaufmerksamkeit* (pp. 9-48). Innsbruck, Wien, Bozen: Studien Verlag.

Prediger, S., & Krägeloh, N. (2015). Low achieving eighth graders learn to crack word problems: a design research project for aligning a strategic scaffolding tool to students' mental processes. *ZDM Mathematics Education*, *47*(6), 947–962. https://doi.org/10.1007/s11858-015-0702-7

Prediger, S., & Zindel, C. (2017). School Academic Language Demands for Understanding Functional Relationships: A Design Research Project on the Role of Language in Reading and Learning. *EURASIA Journal of Mathematics, Science and Technology Education, 13*(7b), 4157-4188. https://doi.org/10.12973/eurasia.2017.00804a

Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E., & Benholz, C. (2018). Language proficiency and mathematics achievement – Empirical study of language-induced obstacles in a high stakes test, the central exam ZP10. *Journal für Mathematik-Didaktik*, *39*(Suppl. 1), 1-26. https://doi.org/10.1007/s13138-018-0126-3

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further exami-nation. *Psychonomic Bulletin & Review, 3*, 504-509. https://doi.org/10.3758/BF03214555

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *Psychology of reading*. New York: Psychology Press. https://doi.org/10.4324/9780203155158

Reusser, K. (1989). Vom Text zur Situation zur Gleichung. *Kognitive Simulation von Sprachverständnis und Mathematisierung beim Lösen von Textaufgaben*. Universität Bern: Habilitationsschrift.

Reusser, K. (1993). Tutoring systems and pedagogical theory: representational tools for understanding, plan-ning, and reflection in problem solving. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as Cognitive Tools* (pp. 143-177). Mahwah: Erlbaum.

Reusser, K. (1997). Erwerb mathematische Kompetenzen: Literaturüberblick. In F. Weinert & A. Helmke (Eds.), *Entwicklung im Grundschulalter* (pp. 141-155). Weinheim: Beltz.

Schindler, M., & Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics, 101*(1), 123-139. https://doi.org/10.1007/s10649-019-9878-z

Schindler, M., Lilienthal, A. J., Chadalavada, R., & Ögren, M. (2016). Creativity in the eye of the student. Refin-ing investigations of mathematical creativity using eye-tracking goggles. In C. Csíkos, A. Rausch, & J. Szitányi (Eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education (PME)* (Vol. 4, pp. 163-170). Szeged: PME.

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: a research review. *Reading and Writing Quarterly, 23*(2), 139-159. https://doi.org/10.1080/10573560601158461

Smith, M. S. (2008). Morphological and syntactic awareness in foreign/second language learning. In J. Cenoz & N.H. Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 6, pp. 179-191). Heidelberg: Springer.

Solano-Flores, G. (2010). Function and Form in Research on Language and Mathematics Education. In J. Mos-chkovich (Ed.), *Language and Mathematics Education: Multiple Perspectives and Directions for Research* (pp. 113-149). Charlotte: Information Age.

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. https://doi.org/10.1007/s13138-019-00144-0

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathemat-ics education research: A systematic literature review. *Educational Studies in Mathematics, 104*(2), 147-200. https://doi.org/10.1007/s10649-020-09948-1

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

van der Schoot, M., Bakker Arkema, A., Horsley, T., & van Lieshout, E. (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*, 58-66. https://doi.org/10.1016/j.cedpsych.2008.07.002

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

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

Wildemann, A., Akbulut, M., & Bien-Miller, L. (2016). Mehrsprachige Sprachbewusstheit zum Ende der Grundschulzeit. Vorstellung und Diskussion eines Elizitationsverfahrens. *Zeitschrift für Interkulturellen Fremdsprachenunterricht, 21*(2), 42-56.

Wolf, M. K., & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. *Educational Assessment, 14*(3-4), 139-159. https://doi.org/10.1080/10627190903425883

Wood, D., Bruner, J. S., & Ross, G. (1976). The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x