Contemporary Educational Psychology 34 (2009) 58–66

Contents lists available at ScienceDirect

Contemporary Educational Psychology

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

The consistency effect depends on markedness in less successful but not successful problem solvers: An eye movement study in primary school children

Menno van der Schoot [a,*], Annemieke H. Bakker Arkema [b], Tako M. Horsley [a], Ernest C.D.M. van Lieshout [a]

[a] Department of Special Education, VU University Amsterdam, Faculteit der Psychologie en Pedagogiek, Afdeling Orthopedagogiek, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

[b] IVLOS Institute of Education, University of Utrecht, Postbus 80127, 3508 TC Utrecht, The Netherlands

ARTICLE INFO

Article history:
Available online 9 September 2008

Keywords:
Word problem solving
Consistency effect
Markedness
Eye movements

ABSTRACT

This study examined the effects of consistency (relational term consistent vs. inconsistent with required arithmetic operation) and markedness (relational term unmarked ['more than'] vs. marked ['less than']) on word problem solving in 10–12 years old children differing in problem-solving skill. The results showed that for unmarked word problems, less successful problem solvers showed an effect of consistency on regressive eye movements (longer and more regressions to solution-relevant problem information for inconsistent than consistent word problems) but not on error rate. For marked word problems, they showed the opposite pattern (effects of consistency on error rate, not on regressive eye movements). The conclusion was drawn that, like more successful problem solvers, less successful problem solvers can appeal to a problem-model strategy, but that they do so only when the relational term is unmarked. The results were discussed mainly with respect to the linguistic–semantic aspects of word problem solving.

1. Introduction

Solving an arithmetic word problem requires a combination of different skills and functions, such as working memory and reading comprehension (Muth, 1984; Passolunghi & Pazzaglia, 2005; Passolunghi & Siegel, 2001; Swanson, Cooney, & Brock, 1993). In addition, certain aspects of the word problem itself have been shown to influence problem-solving success (De Corte, Verschaffel, & Pauwels, 1990; Kintsch & Greeno, 1985; Lewis & Mayer, 1987; Pape, 2003; Stern, 1993; Verschaffel, De Corte, & Pauwels, 1992). Among these, consistency (relational term consistent vs. inconsistent with the required arithmetic operation) and markedness (relational term unmarked ['more than'/'times more than'] vs. marked ['less than'/'times less than']) are the most well-established (Arendasy, Sommer, & Ponocny, 2005; Clark, 1969; Hegarty, Mayer, & Green, 1992; Lewis & Mayer, 1987; Verschaffel, 1994; Verschaffel et al., 1992). However, the effects of consistency and markedness have not yet been examined in children differing in problem-solving skill. This is surprising since word problems comprise a significant part of the curriculum in most school mathematics programs, and good and poor problem solvers may require different educational methods used in teaching word problem solving. To help fill the gap in the literature on word problem solving in children, the goal of the present study was to examine the effects of consistency and markedness on two-step compare word problem solving in 10–12 years old less successful and successful problem solvers. Thereby, we adopted an eye movement approach so as to gain more insight in the cognitive processes underlying word problem solving.

1.1. The consistency effect

A two-step compare word problem contains a numerical relation between two variables (Fan, Mueller, & Marini, 1994; Mayer, 1981; Riley & Greeno, 1988). Its first sentence typically includes an assignment statement expressing the value of the first variable (e.g., *At Aldi, a bottle of wine costs 7 euro*). The second sentence contains the relational statement expressing the value of the second variable in relation to the first (e.g., *At Boni, a bottle of wine costs 2 euro more than at the Aldi*). The third sentence then asks a question about the value of some quantity of the second variable (e.g., *If you need to buy 4 bottles of wine, how much will you pay at Boni?*). The solution of such a two-step word problem requires the calculation of the value of the second variable (step 1) and the subsequent multiplication of this value by the quantity given in the third sentence (step 2). The arithmetic operator which is needed to calculate the value of the second variable can be derived from the second sentence in which the relational term states whether the value of the second variable is, for example, 'x more than' or 'x less than' the first variable.

The difficulty of a compare word problem depends, among other things, on whether the key relational term is consistent or inconsistent with the required arithmetic operation (Briars & Larkin, 1984; Hegarty, Mayer, & Monk, 1995; Hegarty et al.,

* Corresponding author. Fax: +31 30205988745.
E-mail address: M.van.der.Schoot@psy.vu.nl (M. van der Schoot).

0361-476X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cedpsych.2008.07.002

1992; Lewis & Mayer, 1987; Pape, 2003; Verschaffel, 1994; Verschaffel et al., 1992). In a consistent word problem, the relational statement (e.g., 'more than') primes the required arithmetic operation (addition). On the other hand, in an inconsistent version of a word problem, the relational keyword (e.g., 'less than') primes an inappropriate arithmetic operation (subtraction when the required operation is addition). Problem solvers typically display a higher error rate (Hegarty et al., 1992; Hegarty et al., 1995; Lewis & Mayer, 1987; Pape, 2003; Verschaffel, 1994; Verschaffel et al., 1992) and longer response times (Hegarty et al., 1992; Verschaffel, 1994; Verschaffel et al., 1992) on inconsistent than consistent word problems. This finding is referred to as the consistency effect (Lewis & Mayer, 1987). Mostly, problem solvers make a so-called reversal error on an inconsistent word problem because they use the keyword-congruent and thus reverse arithmetic operation.

Hegarty et al. (1995) compared the strategic word problem comprehension processes which are used by successful problem solvers (who do not make errors on inconsistent problems) and unsuccessful problem solvers (who make errors on inconsistent problems) as they planned solutions for two-step word problems such as described above. Their results showed that successful problem solvers make more use of the problem-model strategy, and that unsuccessful problem solvers make more use of the direct-translation strategy. That is, successful problem solvers translate the problem statement into a qualitative mental model (or representation) of the base type of situation that is hidden in the problem. On the basis of this mental model, they plan and execute the required arithmetic operations. Conversely, unsuccessful problem solvers employ a more superficial strategy as they restrict themselves to selecting the numbers and relational keywords from the problem and base their solution plan and calculations on these. Clearly, this latter strategy results in (more) reversal errors on inconsistent word problems.

The consistency effect has been attributed to the difference in the order in which the information is presented in consistent and inconsistent word problems. According to this account, problem solvers prefer a certain order in which problem information should be presented. In a consistent word problem, the way in which the value of the second variable (V2) is expressed in relation to the first (V1) complies with the preferred presentation order (e.g., 'At V1, product A costs 5 euro. At V2, product A costs 2 euro more than at V1'). In an inconsistent word problem, the presentation of the terms in the relational statement conflicts with this preference (e.g., 'At V1, product A costs 5 euro. This is 2 euro more than at V2'). The idea is that in order to solve an inconsistent word problem, the order of the information need to be reversed to the preferred format. This idea receives support from studies showing that successful problem solvers retell inconsistent word problems in terms of the preferred order (Pape, 2003; Verschaffel, 1994). Unsuccessful problem solvers, on the other hand, are hypothesized to experience representational difficulties with an inconsistent word problem because, among other things, they have trouble with performing the order reversal (Lewis & Mayer, 1987; Verschaffel, 1994). In the present study, we wished to further investigate this hypothesis in children.

1.2. The markedness principle

Besides consistency, another factor which has been found to affect word problem solving is markedness. Several studies showed that it matters whether a word problem contains a marked relational term like 'less than' or an unmarked relational terms like 'more than' (Hegarty et al., 1992; Lewis & Mayer, 1987; Pape, 2003; Verschaffel et al., 1992). In more general terms, Clark (1969) introduced this effect as the lexical marking principle stating that the marked member of antonymous adjective pairs (such as less in more–less, but also short in tall–short or narrow in wide–narrow) are more difficult to process than the unmarked member (more, tall, wide). Clark explained the marked–unmarked processing difference by arguing that the semantic memory representation of the 'negative' pole of an adjective pair (i.e., the marked member) is more complex than that of the 'positive' pole (i.e., the unmarked member; see also Bierwisch & Lang, 1989; Clark & Clark, 1977). The higher semantic complexity of marked than unmarked terms has been demonstrated in memory tasks (less accurate recall of marked terms; Clark & Card, 1969), but also in naming tasks (slower naming responses for marked terms; Schriefers, 1990) and reasoning problems (slower solution times for problems with marked adjectives; French, 1979; Jones, 1970).

Of relevance here is that the markedness principle has also been found to influence the solving of word problems in which the respective marked and unmarked relational terms 'less than' and 'more than' appear. For example, Hegarty et al. (1992) demonstrated that high-accurate students take more time to solve word problems containing marked than unmarked terms, and Verschaffel et al. (1992); experiment 2 and 3), Lewis and Mayer (1987) and Pape (2003) showed that markedness interacts with consistency in such a way that problem solvers find it especially difficult to reverse an inconsistent relational sentence when the relational term is marked (i.e., to solve a problem in the inconsistent-marked condition). These interacting effects of consistency and markedness on word problem solving have been explained by assuming that, due to its semantic complexity, problem solvers are more likely to resist reversing the marked inconsistent relational sentence to the preferred format (Lewis & Mayer, 1987).

1.3. The present study

In the present study, we examined the effects of consistency and markedness on word problem solving in children differing in problem-solving skill. The children came from normal primary schools (grade 5 and 6) and were classified as more and less successful problem solvers on the basis of their performance on two-step word problems. We used two-step word problems to inflate the effects of consistency since research in university students suggested that one-step word problems (in which only the value of the second variable needs to be calculated) may be too easy to evoke the consistency effect (Verschaffel et al., 1992).

We hypothesized that the consistency effect will be more pronounced for less successful than more successful problem solvers. This hypothesis is based on the supposition that more and less successful problem solvers differ in their word problem comprehension strategies. Such differences have been demonstrated in undergraduate students (Hegarty et al., 1995) but it is as yet unclear whether they also hold for primary school children. Specifically, we expected that more successful problem solvers will construct a qualitative mental model of the situation described in a problem, and that less successful problem solvers will selectively focus on the numbers and relational terms in a problem after which they will directly translate these into a set of computations. As explicated before, direct-translation leads to incorrect answers (i.e., reversal errors) in an inconsistent word problem in which the relational term primes an inappropriate arithmetic operation and in which information elements other than the numbers and relational terms are necessary for the solution of the problem.

Following the logic described earlier, we also hypothesized that the consistency effect will be enhanced in marked compared to unmarked word problems. In addition, we explored whether these interacting effects of consistency and markedness would be more manifest for less successful than for more successful problem solvers. Possibly, the less successful problem solvers' representational difficulties with an inconsistent word problem may be inflated by difficulties they might have with dealing with the markedness constraints.

1.4. The role of working memory and reading comprehension in word problem solving

Previous studies identified a number of more basic cognitive processes that underlie individual differences in arithmetic word problem solving. In particular, word problem-solving ability has been found to be related to working memory (e.g., Andersson, 2007) and reading comprehension (e.g., Swanson et al., 1993). To be able to examine group differences adjusted for differences in working memory capacity and reading comprehension level, we included these variables as covariates in the present study.

The involvement of working memory in word problem solving is obvious since word problem solving requires the (coordination of) concurrent processing and storage of information. More specifically, word problem solving is shown to depend on the ability to select, maintain and update solution-relevant, and inhibit solution-irrelevant information (Passolunghi, Cornoldi, & De Liberto, 1999; Passolunghi & Pazzaglia, 2005). Low working memory capacity may thus partly explain why less successful problem solvers tend to resort to the direct-translation strategy (and make reversal errors), since this problem-solving approach makes minimal demands on working memory (Hegarty et al., 1995; Kintsch & Greeno, 1985). Controlling for working memory is therefore expected to somewhat decrease the magnitude of the consistency effect in less successful problem solvers, thereby reducing the above-hypothesized differences with the successful problem solvers.

A number of researchers have also demonstrated a contribution of reading comprehension to arithmetic word problem solving (e.g., Cummins, Kintsch, Reusser, & Weimer, 1988; De Corte et al., 1990; Muth, 1984). Most of them have found effects of variations in the linguistic and semantic aspects of word problems on problem representation and problem-solving performance. A plausible interpretation of these effects is that successfully understanding the language of written word problems is an essential first step in modeling the implicit problem structure that specifies the solution-relevant elements of the problem and how these are numerically related to each other (Kail & Hall, 1999). For example, this interpretation explains the finding that rewording the text of word problems so that they more closely reflect the problem structure leads to improvements in problem-solving accuracy (Davis-Dorsey, Ross, & Morrison, 1991; De Corte & Verschaffel, 1987; De Corte, Verschaffel, & De Win, 1985). Along the same line of thought, it has been argued that comprehension skill may be a factor in mapping mathematical terms onto mathematical operations, especially when the mathematical terms translate into non-obvious operations such as the inconsistent relational statements used in the present study (Lee, Ng, Ng, & Lim, 2004). We therefore predicted that controlling for reading comprehension would reduce the magnitude of the consistency effect, and that it would do so especially for word problems containing the semantically more complex marked relational terms. Assuming that less successful problem solvers are poorer comprehenders than successful problem solvers, the overall effect of reading comprehension skill would presumably be a weakening of the interaction between consistency, markedness and comprehension group.

1.5. Eye movement approach

To gain more insight in the cognitive and strategic processes underlying word problem solving in more and less successful problem solvers, we registered the participants' eye movements during task performance. The eye tracking technique has been typically adopted to the study of reading (comprehension) processes (e.g., Hyönä, Lorch, & Rinck, 2003; Van der Schoot, Vasbinder, Horsley, Reijntjes, & Van Lieshout, in press; Van der Schoot, Vasbinder, Horsley, & Van Lieshout, 2008). The assumptions made by these studies are straightforward: fixation location reflects attention (the word that is looked at apparently has the reader's attention) and fixation duration reflects processing difficulty and/or amount of invested attention (the longer a word is fixated, the more complex it is and/or the deeper it is processed). However, the eye tracking technique has also proven suitable to examine arithmetic word problem-solving processes (e.g., Hegarty et al., 1992; Hegarty et al., 1995). In essence, the assumptions made by the studies on reading are also made by studies on arithmetic word problem solving: the elements in a word problem—whether they are numbers, relational terms, variable names or other cues—that are fixated longer are assumed to be more deeply processed and, consequently, central to the construction of the representation and production of the solution (De Corte et al., 1990; Just & Carpenter, 1980; Rayner & McConkie, 1976; Verschaffel et al., 1992). In line with the findings in undergraduate students (Hegarty et al., 1992; Hegarty et al., 1995), we argued that the use of the different supposed comprehension strategies for word problems will not only be reflected in the children's (reversal) error rate, but also in their eye movement/fixation patterns: the less successful problem solvers using a direct-translation strategy (leading to errors on inconsistent word problems) will focus more on, that is, look longer at, the numbers and relational keywords, whereas the more successful problem solvers using a problem-model strategy will distribute their attention more proportionally over the word problem, paying also attention to other words in the problem that are necessary for constructing the problem model such as the names of the variables in the problem.

We assumed that the above-hypothesized effects of consistency, markedness and problem-solving group on word problem solving will be present in the re-reading times (i.e., the number and duration of regressive eye movements) rather than in the first-pass reading times. This assumption is based on the idea that problem solvers first have to read the problem before they can start solving it (Hegarty et al., 1992; Kintsch & Greeno, 1985). In support of this idea, several studies have demonstrated that the consistency effect occurs during the phase after the initial reading (e.g., De Corte et al., 1990; Hegarty et al., 1992; Hegarty et al., 1995).

2. Method

2.1. Participants

Sixty participants were recruited from a normal primary school (grade 5 and 6). They were native speakers of Dutch and 10–12 years of age ($M = 11.50$, $SD = 0.81$). All children were healthy and had normal or corrected-to-normal vision, and their IQ scores were in the normal range. Participants were excluded from the study if school records revealed (diagnosed) behavioral problems or learning disabilities. Participation was voluntary and parental consent was required. After the experiment, children received a small present.

2.2. Classification of the more and less successful problem-solving group

From the initial sample of 60 children, 20 more successful and 20 less successful problem solvers were selected to participate in the experiment. Selection was based on their performance on the word problems (see below). The 20 participants with the lowest error rate were classified as more successful problem solvers (age: $M = 11.25$, $SD = .72$, 9 boys) and the 20 participants with the highest error rate were classified as less successful problem solvers (age: $M = 11.65$, $SD = .75$, 9 boys). The digit span subtest of the Dutch version of the Wechsler Intelligence Scale for Children—R (see below) revealed that the more successful problem solvers had more working memory capacity than the less successful problem solvers ($t(38) = 4.63$, $p < .001$, $d = 1.46$). In addition, the standardized Test for Reading Comprehension of the Dutch National Institute for Educational Measurement ('CITO Toetsen Begrijpend Lezen', Staphorsius & Krom, 1998, see below) demonstrated that the more successful problem solvers had a higher level of reading comprehension skill than the less successful problem solvers ($t(38) = 6.80$, $p < .001$, $d = .99$).

2.3. Materials and design

Each two-step word problem was presented in Dutch and consisted of three sentences. The first sentence was an assignment statement expressing the value of the first variable, that is, the price of a product at a well-known Dutch store or supermarket (e.g., *At Edah, a box of candles costs 3 euro*). The second sentence contained a relational statement expressing the value of the second variable, that is, the price of this product at another store or supermarket, in relation to the first (e.g., *At Spar, a box of candles costs 2 euro more than at Edah*). In the third sentence, the problem solver is asked to find a multiple of the value of the second variable (e.g., *If you need to buy 6 boxes of candles, how much will you pay at Spar?*). The answer to the word problems always involved first computing the value of the second variable (e.g., $3 + 2 = 5$) and then multiplying the result of this computation by the quantity given in the third sentence (e.g., $6 \times 5 = 30$).

The first step computation of the value of the second variable using the value of the first variable requires an addition (as in the present example), subtraction, multiplication or division operation. The relational terms requiring addition or subtraction were expressed as 'more than' or 'less than', and the relational terms requiring multiplication or division were expressed as 'times more than' or 'times less than'. As will become clear below, the actual choice of the arithmetic operation to perform depends on the experimental condition.

The two problem-solving groups performed in four within-subject conditions formed by the crossing of markedness and consistency. Markedness refers to whether the relational term in the second sentence is either 'more than' or 'times more than' (unmarked), or 'less than' or 'times less than' (marked). Consistency refers to whether the relational term in the second sentence is consistent or inconsistent with the required arithmetic operation. A consistent sentence explicitly expresses the value of the second variable (V2) in relation to the first variable (V1) introduced in the prior sentence (*At V2, product A costs N [euro more/times more/euro less/times less] than at V1*). An inconsistent sentence relates the value of the second variable to the first by using a pronominal reference (*This is N [euro more/times more/euro less /times less] than at V2*). Consequently, the relational term in a consistent word problem primes the appropriate arithmetic operation ('more than' when the required operation is addition, 'times more than' when the required operation is multiplication, 'less than' when the required operation is subtraction, and 'times less than' when the required operation is division), and the relational term in an inconsistent word problem primes the inappropriate arithmetic operation ('more than' when the required operation is subtraction, 'times more than' when the required operation is division, 'less than' when the required operation is addition, and 'times less than' when the required operation is multiplication).

An example of a word problem in each condition is given in Table 1 (for addition/subtraction).

Each participant was presented with eight word problems, two in each condition, one of which required an addition or subtraction operation as the first step, and the other requiring a multiplication or division operation. This logically implies that participants had to perform an addition and multiplication operation in both the unmarked-consistent and marked-inconsistent condition, and a subtraction and division operation in both the unmarked-inconsistent and marked-consistent condition. This coupling of conditions and arithmetic operations is inherent to the intrinsic structure of word problems differing in their consistency and lexical marking, and cannot be avoided in a full factorial design in which the two levels of markedness are crossed with the two levels of consistency (see also Hegarty et al., 1992; Hegarty et al., 1995; Lewis & Mayer, 1987; Pape, 2003; Verschaffel et al., 1992).

The stimuli were arranged in four material sets, each containing the eight word problems. The order in which the word problems were presented in each set was pseudorandomized. Each set was presented to 15 participants. Across sets and across participants, each word problem occurred equally often in the unmarked/consistent, unmarked/inconsistent, marked/consistent and marked/inconsistent version to ensure full combination of conditions and materials. Across word problems, we controlled for the difficulty of the required calculations, and for the number of letters in the names of the variables (i.e., stores) and products.

The numerical values used in the word problems were selected on basis of the following rules: (1) The answers of the first step of the operation were below 10, (2) The final answers were between the 14 and 40, (3) None of the first step or final answers contained a fraction of a number or negative number, (4) No numerical value occurred twice in the same problem, and (5) None of the (possible) answers resulted in 1. The numerical values used in the unmarked/consistent and marked/inconsistent problems (requiring addition and multiplication) and the numerical values used in the unmarked/inconsistent and marked/consistent problems (requiring subtraction and division) were matched for magnitude.

Two-step word problems instead of one-step problems were administered to avoid a possible ceiling effect (Hegarty et al., 1992). Verschaffel et al. (1992) concluded in their first experiment that the consistency effect would only occur when the problems are of a certain difficulty level. Therefore, we composed the word problems in line with those of Lewis and Mayer (1987).

Table 1
Example of word problem in the four conditions formed by the crossing of consistency (consistent vs. inconsistent) and markedness (unmarked vs. marked)

	Consistent	Inconsistent
Unmarked	At Aldi, a bottle of wine costs 7 euro *At Boni, a bottle of wine costs 2 euro* ***more than*** *at the Aldi* If you need to buy 4 bottles of wine, how much will you pay at Boni? (Answer: $7 + 2 = 9$, $4 \times 9 = 36$)	At Aldi, a bottle of wine costs 7 euro *This is 2 euro* ***more than*** *at Boni* If you need to buy 4 bottles of wine, how much will you pay at Boni? (Answer: $7 - 2 = 5$, $4 \times 5 = 20$)
Marked	At Aldi, a bottle of wine costs 7 euro *At Boni, a bottle of wine costs 2 euro* ***less than*** *at the Aldi* If you need to buy 4 bottles of wine, how much will you pay at Boni? (Answer: $7 - 2 = 5$, $4 \times 5 = 20$)	At Aldi, a bottle of wine costs 7 euro *This is 2 euro* ***less than*** *at Boni* If you need to buy 4 bottles of wine, how much will you pay at Boni? (Answer: $7 + 2 = 9$, $4 \times 9 = 36$)

Note: The examples given concern addition and subtraction. For multiplication or division, the relational terms are expressed as 'times more than' or 'times less than'.

2.4. Individual difference measures

To assess children's working memory capacity, the digit span subtest of the Dutch version of the Wechsler Intelligence Scale for Children—R (Van Haasen, 1986) was administered. In the forward digit span task, the participant was asked to repeat series of an increasing number of digits (starting with two digits, ending with eight digits). In the backward digit span task, the participant was required to recall series of an increasing number of digits in the reverse order of presentation (starting with two, ending with eight digits). The tasks ended after the participant had incorrectly repeated two series with the same number of digits.

Reading comprehension level was assessed by the (grade 5 and grade 6 versions of the) standardized Test for Reading Comprehension of the Dutch National Institute for Educational Measurement ('Cito Toetsen Begrijpend Lezen', Staphorsius & Krom, 1998). This test is part of the standard Dutch CITO pupil monitoring system and is designed to determine general reading comprehension level in primary school children. Each test contains three modules, each consisting of a (narrative or expository) text and 25 multiple-choice questions. The questions pertain to either the word, sentence or text level and tap both the textbase and situational representation that the reader constructed from the text (e.g., Kintsch, 1988).

2.5. Apparatus

Eye movements/fixations were recorded during the solving of the word problems with the EYELINK® II eye-tracker, an infrared video-based tracking system manufactured by SR Research Ltd. (Mississauga, Canada). The EYELINK® II consists of three miniature cameras mounted on a headband. Two high-speed cameras (CCD sensors) with built-in infrared illuminators allow binocular eye tracking. An optical high-speed camera integrated into the headband monitors head movements so that point of gaze can be accurately tracked even when a participant moves its head. The headband weighs 450 g in total. The system uses a corneal reflection method in combination with pupil tracking, permitting stable tracking of eye position regardless of muscle tremor, environmental vibration, or headband slippage.

The eye-tracker uses an online saccade parser to generate saccade and fixation events. Motion (0.1°), velocity (30°/s), and acceleration ($8000°/s^2$) thresholds are used to determine the start and end of saccades. When the eye movement goes above the thresholds a saccade start is identified, when the eye movement goes below the thresholds a saccade end is identified. The period between saccades is identified as a fixation.

The best eye to record was automatically selected during a calibration procedure. Calibration was accepted when the worst error in gaze position was smaller than 1.5° and average error was smaller than 1.0°. The cameras sampled pupil location at the rate of 250 Hz. Participants in the experiment were seated such that the distance between the monitor (51 cm diagonal with a resolution of 1025×768) on which the word problems were displayed and their eyes was approximately 70 cm. At this distance, three letters of the presented word problem subtended 1° of visual angle.

2.6. Procedure

Before the start of the experimental task, children were informed that the study was designed to examine the solving of word problems displayed on a screen. A short calibration procedure and a practice session of one word problem preceded the administration of the word problems to accustom the children to the eye tracking equipment.

The word problems were presented on the computer screen one at a time. The first two sentences each consisted of one line of text, the third sentence was presented in two lines. The experimenter presented the next word problem as soon as the child verbally stated the answer to the current one. The experimental word problems were interspersed with four filler problems to prevent the participants from becoming aware of the purpose of the experiment. The filler problems were word problems with a different structure. The order in which the filler problems were presented was pseudorandomized with the exception that a filler problem was presented after at least every two experimental word problems.

Participants were given a break after receiving half of the word problems. During the break, the eye tracking equipment was recalibrated. After the word problem-solving task, the tests for working memory and reading comprehension were administered (since the Test for Reading Comprehension is part of the standard Dutch CITO pupil monitoring system, the recent scores on this test could be obtained from the school records). All tests instructions were standardized and were read aloud by the experimenter. In total, the experiment lasted approximately 45 min.

2.7. Data analysis

Following Hegarty et al. (1995), regression times for three regions in each word problem were calculated: the numbers, the relational terms and the variable names. In addition, we counted the number of times each participant looked back at numbers, relational terms and variable names (i.e., regression run count). These three regions are considered as most important regions for solving word problems correctly, and enabled us to make a distinction between the direct-translation strategy and the problem-model strategy (Hegarty et al., 1995). According to Hegarty et al., less successful problems solvers apply a direct-translation strategy and focus on the numbers and relational term of a word problem. More successful problem solvers on the other hand, adopt a problem-model strategy and focus on the variable names in addition to the numbers and relational term.

We analyzed the participants' regression times and regression run counts because additional processing for inconsistent word 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). Regression time is defined as the total fixation time on a region minus the first pass time (Frisson & Pickering, 1999; Rayner, 1998).

3. Results

Table 2 presents the means and standard deviations of percentage of errors, regression time and regression run count as a function of consistency (consistent vs. inconsistent), markedness (marked vs. unmarked) and group (successful vs. less successful problem solvers). On the proportion of errors, regression time and regression run count, an overall $2 \times 2 \times 2$ analysis of variance (ANOVA) was conducted with Consistency and Markedness as within-subject variables and Group as between-subject variable. The analyses on regression time and regression run count were performed separately for (1) numbers and relational terms, (2) variable names, and (3) numbers, relational terms and variable names (i.e., total regression time/regression run count).

3.1. Error rate analysis

Table 2 shows that overall, less successful problem solvers displayed a higher error rate on the word problems than successful

Table 2
Means and standard deviations of percentage of errors, regression time and regression run count as a function of consistency (consistent vs. inconsistent), markedness (marked vs. unmarked) and group (more successful vs. less successful problem solvers)

	More successful problem solvers								Less successful problem solvers							
	Consistent				Inconsistent				Consistent				Inconsistent			
	Marked		Unmarked		Marked		Unmarked		Marked		Unmarked		Marked		Unmarked	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Percentage of errors	0	0	2.5	11.2	20	34.0	12.5	22.2	55	27.6	62.5	31.9	97.5	11.2	65.0	36.6
Regression time (ms)																
Total	1132.3	709.8	1220.0	788.5	1706.0	982.0	1411.2	1028.4	2033.1	1456.5	1932.3	1037.9	2292.1	1563.0	2917.0	1606.0
Numbers/relational terms	1431.0	970.1	1458.4	1105.8	2083.6	1303.2	1686.4	1333.4	2390.0	1691.1	2328.8	1234.5	2692.5	1804.8	3520.9	2027.3
Variable names	534.7	358.2	743.3	429.0	1040.8	707.9	860.8	587.6	1319.2	1161.4	1139.5	784.1	1491.1	1368.5	1709.2	979.1
Regression run count																
Total	2.6	1.2	3.1	1.4	3.9	1.7	3.3	1.9	4.8	3.3	4.5	2.5	5.3	3.6	6.8	3.4
Numbers/relational terms	3.2	1.6	3.5	1.8	4.6	2.1	3.8	2.3	5.8	3.9	5.2	2.9	6.4	4.2	8.2	4.3
Variable names	1.4	.8	2.2	1.1	2.6	1.5	2.2	1.2	3.1	2.4	2.9	1.9	3.0	2.7	4.0	2.1

Note: Total = mean regression time and mean regression run count for numbers, relational terms and variable names.

problem solvers ($F(1,38) = 222.53$, $p < .0005$, MSE = .27, $\eta_p^2 = .85$; 0.01 is considered a small partial eta-squared effect size, 0.06 is considered a medium effect, and 0.14 is considered a large effect [Stevens, 2002]). Note that this effect is not of particular interest since it merely reflects the selection procedure. Further, more and less successful problem solvers made more errors on inconsistent than consistent word problems ($F(1,38) = 22.15$, $p < .001$, MSE = .25, $\eta_p^2 = .36$). In less successful problem solvers, this consistency effect was pronounced for marked but absent for unmarked word problems (Consistency × Markedness interaction: $F(1,19) = 6.91$, $p < .05$, MSE = .46, $\eta_p^2 = .28$). In more successful problem solvers, the consistency effect was larger for marked than unmarked word problems as well (Consistency × Markedness interaction: $F(1,19) = 1.65$, *ns*), but the difference was considerably less pronounced than in less successful problem solvers. The Consistency × Markedness × Group interaction approached significance at the 0.05 level ($F(1,38) = 3.08$, $p < .08$, MSE = .29, $\eta_p^2 = .08$), as did the interaction when working memory was entered as a covariate ($F(1,37) = 4.03$, $p = .052$, MSE = .29, $\eta_p^2 = .098$). When reading comprehension skill was used as the covariate, the three-way interaction was clearly not significant ($F(1,37) = .98$, *ns*).

3.2. Regression time analysis

The analyses on regression time when calculated *separately* for numbers/relational terms and variable names each yielded a pattern of results comparable with the below analysis on the *total* regression time on numbers, relational terms and variable names. This was evidenced by the finding that the Consistency × Markedness × Group interaction did not vary as a function of Information Type (numbers/relational terms vs. variable names) ($F(1,38) = 0.52$, *ns*). Therefore, only the total regression time data will be reported here. The same logic applies to the analyses on the regression run count (see below; $F(1,38) = 1.60$, *ns*).

From Table 2, it can be seen that the pattern of regression times coincide with the pattern of error rate. Overall, inconsistent word problems yielded longer regression times than consistent word problems ($F(1,38) = 11.29$, $p < .01$, MSE = 893299.90, $\eta_p^2 = .23$). In less successful problem solvers, the consistency effect depended on Markedness: the inconsistent − consistent difference in regression time was small for marked but pronounced for unmarked word problems. In this group, inconsistent word problems thus provoked longer regressive fixation times than consistent word problems as long as the problems did not contain marked relational terms. In more successful problem solvers, the consistency effect did not vary as a function of Markedness, at least not to the degree as in less successful problem solvers. The above pattern of results was partly evidenced by the Consistency × Markedness × Group interaction which approached significance at the 0.05 level without controlling for working memory ($F(1,38)$ 3.46, $p < .07$, MSE = 888380.11, $\eta_p^2 = .08$), but which reached conventional levels of significance once working memory was entered as covariate ($F(1,37) = 4.41$, $p < .05$, MSE = 887376.11, $\eta_p^2 = .11$). When reading comprehension skill was used as the covariate, the three-way interaction was clearly not significant ($F(1,37) = .004$, *ns*).

3.3. Regression run count analysis

From Table 2, it can be seen that the pattern of regression run count corresponds with the patterns of regression time and error rate. Overall, participants made more regressions on inconsistent than consistent word problems ($F(1,38) = 13.50$, $p < .001$, MSE = 3.48, $\eta_p^2 = .26$). In less successful problem solvers, the consistency effect again depended on Markedness: the inconsistent – consistent difference in regression run count was almost absent for marked but pronounced for unmarked word problems (Consistency × Markedness interaction: $F(1,19) = 4.31$, $p = .052$, MSE = 4.45, $\eta_p^2 = .19$). In this group, inconsistent word problems thus gave rise to more regressions than consistent word problems but only when the problems contained unmarked relational terms. In more successful problem solvers, the consistency effect did not vary as a function of Markedness ($F(1,19) = 2.25$, *ns*). The above pattern of results was reflected in the Consistency × Markedness × Group interaction ($F(1,38) = 6.60$, $p < .05$, MSE = 3.54, $\eta_p^2 = .15$). This three-way interaction remained significant after controlling for working memory ($F(1,37) = 7.33$, $p < .05$, MSE = 3.52, $\eta_p^2 = .17$), but disappeared when reading comprehension skill was entered as the covariate ($F(1,37) = 0.5$, *ns*).

4. Discussion

This study investigated the effects of consistency and markedness on word problem solving in 10–12 years old successful and less successful problem solvers. We hypothesized that the effect of consistency would be enhanced in marked compared to unmarked word problems, and that this consistency by markedness interaction might be more manifest in less than more successful problem solvers. The error rate analysis confirmed this hypothesis. The consistency effect appeared to depend on markedness in less successful but not (or much less so) in successful problem solvers. In less successful problem solvers, the consistency effect was pronounced for marked and absent for unmarked word problems. In successful problem solvers, the consistency effect was only slightly larger for marked than unmarked word problems (for which it was practically absent).

Interestingly, the analyses on both the duration and the number of regressions coincided with the error rate pattern. In less successful problem solvers, the consistency effect again varied as a function of markedness: inconsistent – consistent differences in regression time and regression run count were respectively small and absent for marked but pronounced for unmarked word problems. Apparently, inconsistent word problems provoked less successful problem solvers to make longer and more regressions than consistent word problems as long as the problems contained unmarked relational terms. If they contained marked relational terms, less successful problem solvers seemed no longer inclined to look back at the solution-relevant regions (numbers, relational terms and variable names) in the word problem. In more successful problem solvers, the consistency effect did not depend on markedness, at least not to the degree as in less successful problem solvers.

What do the findings tell us about the strategic word problem-solving processes in less successful problem solvers (hypothesized to make errors on inconsistent problems) and more successful problem solvers (hypothesized to make no or less errors on inconsistent problems)? Earlier, we explained that the direct-translation strategy is destined to result in (reversal) errors on inconsistent word problems while the problem-model strategy does not (Briars & Larkin, 1984; Hegarty et al., 1995). In this light, the conclusion that must be drawn from the error rate pattern is straightforward: unlike earlier suggestions (Hegarty et al., 1992; Hegarty et al., 1995; Verschaffel, 1994), not only more successful problem solvers but also less successful problem solvers can standardly appeal to a problem-model strategy. However, the latter only do so when the key relational term is unmarked. After all, for unmarked word problems, less successful problem solvers did not show the consistency effect, and only the use of a problem-model strategy could have prevented them from making (more) reversal errors on the inconsistent problem versions. The use of the problem-model strategy in the unmarked condition was reflected in their regressive eye movement pattern in two ways. First, unmarked inconsistent word problems induced longer and more regressions to the numbers, relational terms and variable names than unmarked consistent word problems. Second, less successful problem solvers did not show an attentional bias towards the numbers and relational term (which is typical for a direct-translation strategy) since the above error rate pattern was also observed for the variable names.

For marked word problems, less successful problem solvers showed the opposite pattern. Here, the presence of the consistency effect at the performance level (i.e., more errors on inconsistent than consistent word problems) was accompanied (and partly accounted for) by the absence of an effect of consistency at the attentional level (equal durations and numbers of regressions on inconsistent and consistent word problems). So this leaves us with the question why less successful problem solvers seem to deploy a problem-model strategy for unmarked but not marked word problems. Or, stated differently: why can less successful problem solvers convert the relational term ‘more than’ to a subtraction operation, but not the relational term ‘less than’ to an addition operation? Probably, the answer to this question resides in the higher semantic complexity of the marked relational term ‘less than’ (Clark, 1969; Clark & Clark, 1977; Schriefers, 1990). Lewis and Mayer (1987) already suggested that, due to its semantic complexity, problem solvers are more likely to resist reversing a marked than an unmarked inconsistent word problem. The present study shows that in 10–12 years old children, this seems to be true especially for the less successful problem solvers.

Based on the assumption that overcoming the semantic complexity of a marked relational term may draw on one or more component skills of reading comprehension, we predicted that after statistically controlling for general reading comprehension skill, the effects described above would weaken or even disappear. This prediction proved to be correct. When reading comprehension skill was used as a covariate, the interacting effects of consistency, markedness and problem-solving group on error rate, regression time and regression run count dissipated. Thus, processing marked relational terms seems to also rely on more general reading comprehension skill. To understand this, remember that Clark (1969) explained the markedness principle by assuming that the semantic memory representation of a marked member of an antonymous adjective pair (less in more–less, or short in tall–short) is more complex than the semantic memory representation of the unmarked member (more, tall). This difference in semantic complexity between marked and unmarked terms may be related (at least in part) to a difference in their frequency of occurrence (Schriefers, 1990). Whereas the marked member of an antonymous adjective pair is used only in its contrastive, ‘negative’ sense (‘Peter has *less* marbles than David’), the unmarked member is used in two senses: the contrastive, ‘positive’ sense (‘Peter has *more* marbles than David’) but also a neutral, nominal sense (‘Does she have *more* than one child?’, ‘How *tall* is he?’) (e.g., French, 1979; Goodwin & Johnson-Laird, 2005). As for word problem solving, this may implicate that the memory representation of a marked relational statement is more ‘fixed’ than the memory representation of an unmarked relational statement. Possibly, the fixedness of a marked relational statement’s memory representation may hinder the less successful problem solvers’ ability to reverse the relational statement in the inconsistent condition (in which it primes the inappropriate arithmetic operation). At least, the present findings suggest that it may hinder them if they are deficient in some comprehension-related skill which helps to perform the statement reversal. Interestingly, a similar suggestion has been made by Lee et al. (2004) who claimed that comprehension abilities may play a role in the mapping of mathematical terms onto mathematical operations, especially when the mathematical terms translate into non-obvious operations (as is the case here). However, future research should determine the exact nature of these abilities. In any case, there does not seem to be a role of working memory in reversing marked inconsistent word problems. This conclusion is based on the finding that working memory did not weaken the interacting effects of consistency, markedness and problem-solving group on both error rate, regression time and regression run count.

To conclude, this study shows that less successful problem solvers are capable of applying an effective problem-model strategy for unmarked word problems (effects of consistency on regressive eye movements, not on error rate) but not marked word problems (effects of consistency on error rate, not on regressive eye movements), and that, possibly, this difference may be attributed to a deficiency in some comprehension-related skill required to reverse a marked inconsistent relational sentence. The relevance of this conclusion resides in its implications for instruction and intervention. In general, the goals are to teach the strategic word problem-solving processes that have been found to be important for problem-solving success in laboratory experiments such as this one, and, at the individual level, to remediate deficiencies in these processes. The main implication of the present study is that educational methods should not only emphasize the strategic aspects of word problem solving (how to translate the problem statement into an effective mental model?) but also the more linguistic aspects (how to process marked relational terms appearing in the problem statement?).

4.1. Limitations and implications for future research

Finally, we point to some methodological and theoretical issues that should be taken into consideration in further research.

First, it should be acknowledged that error rate in word problem solving was used to create a more and less successful group, but also to measure performance on the experimental task. As indicated previously, no significance could therefore be attached to the main effect of group on error rate since it merely reflects the selection procedure. However, the goal of this study was to find out whether differences between more and less successful word problem solvers reflect not just differences in the general performance level, but also differences that are brought about by specific performance patterns, that is, a greater weakness of especially the less successful problem solvers to deal with (the combination of) inconsistency and markedness. Group specific consistency by markedness interactions were indeed observed in the present study, and, importantly, the particular error patterns in both groups did not necessarily follow from the group difference in general performance level.

Second, we should take a critical look at the coupling of experimental condition and required arithmetic operation in experiments using word problems such the present ones. As indicated previously, the four conditions that result from the combination of consistency and markedness cannot be independently varied according to an arithmetic operation and its inverse. That is, in the domain of addition and subtraction, unmarked-consistent and marked-inconsistent problems always concern addition, whereas the two other problem types (unmarked-inconsistent and marked-consistent) always concern subtraction. The same logical constraint on the choice of the operation holds for multiplication (unmarked-consistent and marked-inconsistent conditions) and division (unmarked-inconsistent and marked-consistent conditions). As for the present experimental set-up, this implies that participants had to perform one addition and one multiplication operation in both the unmarked-consistent and marked-inconsistent condition, and one subtraction and one division operation in both the unmarked-inconsistent and marked-consistent condition. This type of coupling of conditions and operations is inevitable when using word problems differing in their consistency and lexical marking (see also Hegarty et al., 1992; Hegarty et al., 1995; Lewis & Mayer, 1987; Pape, 2003; Verschaffel et al., 1992). Clearly, this raises the question whether, in the less successful problem solvers, operation difficulty may have inflated the difference in the percentage of errors between the marked-inconsistent and unmarked-inconsistent condition, and the marked-inconsistent and marked-consistent condition. In other words, can the higher number of errors in the marked-inconsistent condition be (partly) attributed to the necessity to add or multiply, and can the lower number of errors in the unmarked-inconsistent and marked-consistent condition be (partly) attributed to the necessity to subtract or divide? The literature is clear on this matter: in terms of error rate, addition is less difficult than subtraction, and multiplication is less difficult than division (e.g., Campbell, Fuchs-Lacelle, & Phenix, 2006). This means that if difficulty of arithmetic operation was a factor in the results, it would have worked against the obtained results rather than that it helped to produce them. As far as the reliability of the results is concerned, we can thus conclude that the coupling of conditions and arithmetic operations does not constitute a significant confounder in this study. At least, it does not undermine the above provided explanation of the results in terms of the interacting effects of markedness and consistency.

Third, it should be noted that some of the effect sizes in this study were small. Partly, this may be due to a ceiling effect in the inconsistent/marked condition in which the less successful problem solvers displayed a 97.5% error rate. This almost lowest-level of performance may have put an artificial limit on the effect sizes. Probably, the provided explanation of the consistency $\times$ markedness interaction in this group would have received even more empirical support without this constraint. In future research, somewhat more difficult word problems should be used to prevent a ceiling effect.

Finally, it should be acknowledged that there may have been a difference in practice-related performance improvements in more and less successful problem solvers. According to Haider and Frensch's (1996) information–reduction hypothesis, people learn, with practice, to distinguish between task-relevant and task-irrelevant information and to limit their processing to task-relevant information. The selective use of information is an important factor in problem solving (e.g., Bransford, Sherwood, Vye, & Rieser, 1986), and it is thought to be the result of a conscious decision to no longer attend to components of the task that contain irrelevant or redundant information (Haider & Frensch, 1999). Of relevance here is that children may differ in the extent to which they benefit from practice with word problem solving. For example, both before and during the present experiment, especially the more successful problem solvers may have increased their knowledge about which type of information in two-step word problems should and should not be processed. Clearly, this may have affected the numbers and durations of their regressive eye fixations on solution-relevant (numbers, relational terms, variable names) and solution-irrelevant information.

References

Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Applied Cognitive psychology, 21*, 1201–1216.

Arendasy, M., Sommer, M., & Ponocny, I. (2005). Psychometric approaches help resolve competing cognitive models: When less is more than it seems. *Cognition and Instruction, 23*, 503–521.

Bierwisch, M., & Lang, E. (1989). Somewhat longer—much deeper—further and further: Epilogue to the dimensional adjective project. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation* (pp. 471–514). Berlin, Heidelberg, New York, Tokyo: Springer-Verlag.

Bransford, J., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching thinking and problem-solving—research foundations. *American Psychologist, 41*, 1078–1089.

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

Campbell, J. I. D., Fuchs-Lacelle, S., & Phenix, T. L. (2006). Identical elements model of arithmetic memory: Extension to addition and subtraction. *Memory & Cognition, 34*, 347–633.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review, 76*, 387–404.

Clark, H. H., & Card, S. K. (1969). Role of semantics in remembering comparative sentences. *Journal of Experimental Psychology, 82*, 545–553.

Clark, H., & Clark, E. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt, Brace, Jovanovich, Inc..

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

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*, 61–68.

De Corte, E., & Verschaffel, L. (1987). The effect of semantic structure on 1st-graders strategies for solving addition and subtraction word problems. *Journal for Research in Mathematics Education, 18*, 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*, 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*, 359–365.

Fan, N., Mueller, J. H., & Marini, A. E. (1994). Solving difference problems—wording primes coordination. *Cognition and Instruction, 12*, 355–369.

French, P. L. (1979). Linguistic marking, strategy, and affect in syllogistic reasoning. *Journal of Psycholinguistic Research, 8*, 425–449.

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology—Learning Memory and Cognition, 25*, 1366–1383.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review, 112*, 468–493.

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology, 30*, 304–337.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information–reduction hypothesis. *Journal of Experimental Psychology—Learning Memory and Cognition, 25*, 172–190.

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology, 84*, 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*, 18–32.

Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science.

Jones, S. (1970). Visual and verbal processes in problem-solving. *Cognitive Psychology, 1*, 201–214.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.

Kail, R., & Hall, L. K. (1999). Sources of developmental change in children's word-problem performance. *Journal of Educational Psychology, 91*, 660–668.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95*, 163–183.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*, 109–129.

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*, 140–158.

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

Mayer, R. E. (1981). Frequency norms and structural-analysis of algebra story problems. *Instructional Science, 10*, 135–175.

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

Pape, S. J. (2003). Compare word problems: Consistency hypothesis revisited. *Contemporary Educational Psychology, 28*, 396–421.

Passolunghi, M. C., Cornoldi, C., & De Liberto, S. (1999). Working memory and intrusions or irrelevant information in a group of specific poor problem solvers. *Memory and Cognition, 27*, 779–790.

Passolunghi, M. C., & Pazzaglia, F. (2005). A comparison of updating processing in children good or poor in arithmetic word problem-solving. *Learning and Individual Differences, 15*, 257–269.

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*, 44–57.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., & McConkie, G. W. (1976). What guides a readers eye movements? *Vision Research, 16*, 829–837.

Riley, S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction, 5*, 49–101.

Schriefers, H. (1990). Lexical and conceptual factors in the naming of relations. *Cognitive Psychology, 22*, 111–142.

Staphorsius, G., & Krom, R. (1998). *Cito Toetsen begrijpend lezen [Reading comprehension test of the Dutch National Institute for Educational Measurement]* (2nd ed.). Arnhem, The Netherlands: Citogroep.

Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *Journal of Educational Psychology, 85*, 7–23.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

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*, 374–395.

Van der Schoot, M., Vasbinder, A. L., Horsley, T. M., Reijntjes, A. & Van Lieshout, E. C. D. M. (in press). Lexical ambiguity resolution in good and poor comprehenders: An eye fixation and self-paced reading study in primary school children. *Journal of Educational Psychology*.

Van der Schoot, M., Vasbinder, A. L., Horsley, T. M., & Van Lieshout, E. C. D. M. (2008). The role of two reading strategies in text comprehension: An eye fixation study in primary school children. *Journal of Research in Reading, 31*, 203–223.

Van Haasen, P. P. (1986). WISC-R; Nederlandstalige uitgave. *Scoring en normen [Wechsler Intelligence Scale for Children – Revised; Dutch version. Scoring and norms]*. Lisse, The Netherlands: Swets & Zeitlinger.

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*, 141–165.

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*, 85–94.