When a half and a half are not a whole: Putting word problems into context

by Cheryll L. Fitzpatrick A Dissertation submitted

to the School of Graduate Studies in partial fulfillment of the

requirements for the degree of

Doctorate of Philosophy

Psychology Department

Memorial University of Newfoundland

May, 2018

St. John's Newfoundland and Labrador

Abstract

Math word problems can be quite challenging to students. They learn early in their formal education that there is a proper "procedural recipe" to follow when solving words problems. Typically, it consists of taking the numbers and keyword(s) that indicate what mathematical operation should be used coupled with whatever lesson they happen to be learning at that moment. These word problems are usually very straightforward, applying one mathematical operation to all the numbers present in the problem, to a situation that largely ignores any inkling of reality, and does not include any extraneous information. This style of mathematical problem solving however is not conducive to applying mathematics in the real world. In the real-world situations are messy, there are often many unknowns, and there may not always be one correct answer. How does children's formal education prepare them for applying their mathematical knowledge to real-world situations? The research examining this topic has consistently shown that students have trouble incorporating their real-world knowledge into their solution process for mathematical word problems that require realistic considerations.

This dissertation investigates the relation between realistic problem solving and general academic skills while also testing procedures meant to improve realistic problem solving in Grade 6 students. After reviewing the literature on this topic in Chapter 1, Chapter 2 focuses on how general academic abilities play a part in children's abilities to use their real-world knowledge. Chapter 3 examines interventions aimed at increasing students realistic responding to realistic word problems. The results of Chapter 2 indicate that general academic abilities are not predictors of success on realistic word problems; however, student's ability to provide realistic responses to realistic word problems was an independent predictor of their performance on standard word problems (e.g., those seen in mathematics classrooms). This suggests that students' general problems solving skills are benefitted by their ability to incorporate realistic reasoning and for this reason realistic reasoning is a skill worth nurturing. The result of Chapter 3 revealed that having students respond to realistic word problems with a response sentence was only helpful for boys and only on a particular type of word problem (Division-with-Remainder; see Experiment 1), using examples hinders students' ability to use their realistic knowledge (Experiment 2), and creating a richer backstory in the word problem also does not increase student's use of realistic knowledge. The experiments conducted in Chapter 3 indicate that students are extremely resistant to including realistic knowledge in their solution process. Chapter 4 provides more general discussion of the findings, and included some additional findings not discussed in Chapters 2 and 3. Future directions in this line of research, the implications of the dissertations findings, and some practical applications were also discussed, but the findings reported here point to the difficulty in teaching children to apply real-world information in problem solving situations and the importance of doing so.

Keywords: word problems, math cognition, intervention, educational psychology, academic abilities

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Darcy Hallett. Without your knowledge and guidance, this project would not have been possible. I would also like to thank the members of my committee, Dr. Catherine Penney and Dr. Toni Doyle. Your input has proved to be invaluable and I thank you for your time and commitment to this endeavor. The amount of time and dedication I have been able to provide to my research is in great deal owed to the funding provided by the Social Sciences and Humanities Research Council. I owe many thanks to all the members of the Research Centre for the Development of Mathematical Cognition for your help with creating questions, data collection, and most of all having to listen to numerous presentations on the same topic! I need to express a great amount of gratitude to the Newfoundland English School District for allowing me to conduct research in elementary schools within the province, and to the schools and students who participate. Without your cooperation and participation this project would never have been possible. Finally, I would like to thank my family and friends for all of your support, especially my husband Neill, for your continued encouragement.

Table of Contents

List of Tables

List of Figures

Chapter 1

Putting Mathematical Word Problems into Context

Mathematical word problems, or story problems, have become the most common kind of problem found in formal education (Jonassen, 2003). Yet, the National Mathematics Advisory Panel (U.S. Department of Education, 2008) reports that word problems are one of the three areas for which students have the poorest preparation. *Math word problems* are short, made-up stories, typically describing a situation assumed to be familiar to the reader (Greer, Verschaffel, & De Corte, 2003) and using quantitative relations between various objects or characters that require a mathematical solution (Martin & Bassok, 2005). There are thought to be two functions of mathematical word problems (WPs): (1) as an exercise to practice basic arithmetic operations and, (2) to engage in mathematical modeling by training students to apply formal mathematical knowledge to real-world situations (Gravemeijer, 1997; Greer et al., 2003; Verschaffel, Greer, & De Corte, 2000).

Successful WP solving and formulation have been studied in a variety of contexts, for example, translation ability (Dark & Benbow, 1990), verbal-logical and visual strategies (Kaizer & Shore, 1995), linguistic knowledge, and knowledge about the schema or “game” of WPs (Verschaffel & De Corte, 1997a), computation ability (Kail & Hall, 1999; Zheng, Swanson, & Marcoulides, 2011), reading comprehension (Bjork & Bowyer-Crane, 2013; Muth & Glynn, 1985; Pimperton & Nation, 2010), and the list goes on. One area in particular that has received wide spread attention in European countries – quite possibly as a result of the dual functionality of WPs – is that of realistic WPs. Too often in an elementary school mathematics classroom, one will find a problem such as the following. ‘Farmer Clem is building a fence around his garden. The garden is 8m wide and 8m long. How many meters of wood does Farmer Clem need to build his fence' Students know better than to ask: 'Is there an opening in the fence for a door?' Only in the 'game of word problems' (De Corte, Verschaffel, & Lasure, 1995; Verschaffel et al., 2000) does a problem like this pass for reality.

A WP is thought to be realistic if the key aspects of the WP are taken into account under the typical conditions for that situation outside of school (Palm, 2008). That means, when the situation is considered as being one in the real world, the student must determine if the numerical response continues to make sense in the described situation. Another facet of realistic WPs is that there may be more than one answer, or that there is not enough information to provide an answer, or that only an approximate answer can be provided. Realistic WPs also may include extraneous information that is not needed in the calculation, which is not typical in standard WPs (Verschaffel et al., 2000).

Elementary and middle school students, undergraduate students, and preservice teachers have a strong tendency to disregard real-world knowledge when solving mathematical WPs that require realistic consideration (Inoue, 2005, 2008; Verschaffel, De Corte, & Borghart, 1997; Verschaffel et al., 2000). As mentioned above, these problems often differ from standard word problems not only in their requirement to take real world knowledge into account, but also because the answers require estimation or an acknowledge of its problematic nature. Still, research participants of all ages have found the realistic WPs that have exact answers to be as difficult as those that do not. Undergraduate students do perform better than elementary and middle school students on realistic WPs (30% vs. 17%, respectively; Inoue, 2005, 2008), but they are still responding in a realistic way less than half of the time. Verschaffel et al. (1997) report that in their sample of 332 preservice teachers, only 48% of all responses were categorized as realistic responses. These results indicate that adults are in fact better at these kinds of problems than elementary and middle school students; however, even at higher levels of education (e.g., preservice teachers) realistic responses are occurring only half of the time.

Ignoring realistic considerations in WPs can lead to students ignoring the content of the problem and reporting nonsensical solutions. The most well-known example of this effect is the ‘Age of the Captain’ problem. This problem was originally given to a group of 1st and 2nd grade students (as cited in Verschaffel et al., 2000) and was presented as follows: “There are 26 sheep and 10 goats on a ship. How old is the captain?” Surprisingly, the majority of children added the number of sheep and goats to provide a solution of 36 as being the age of the captain. Why are so many students attempting to answer such an irrational and unsolvable problem? This ‘suspension of sense making’ (Schoenfeld, 1991; Verschaffel & De Corte, 1997a) has been replicated in countries all over the world (e.g., Belgium, Germany, Japan, China, Northern Ireland, Switzerland, and Venezuela; Reusser & Stebler, 1997; Schoenfeld, 1991; Verschaffel et al., 2000; Xin, Lin, Zhang, & Yan, 2007). Initially, the results were thought to be due to a lack of real-world knowledge possessed by students, specifically that they were not able to properly assess the situational aspects of the problems and therefore did not understand that the problem could not be solved with the given information (Hidalgo, 1997 as cited in Verschaffel et al., 2000). It appears as though students are simply ignoring the story all together, and putting the numbers stated in the problem into an operation that seems appropriate, which, in this case, is addition.

Although Hidalgo's notion may argue that students do not have the real-world knowledge needed for these problems (as cited in Verschaffel et al., 2000), this cannot explain the nonsensical answers provided by students to the 'Age of the Captain' problem. Even young children know that the number of sheep and the number of goats does not add up to somebody's age. Instead, students' ability to comprehend the given text and to assess the situational aspects of the problem may be a better explanation for what is happening. Mathematical WPs are, by definition, computational problems that are described by a story or a situation. It is the job of the student to successfully read, comprehend, and solve the problem accordingly. As previously mentioned, successful WP performance has been examined in terms of reading comprehension (Bjork & Bowyer-Crane, 2013; Muth & Glynn, 1985; Pimperton & Nation, 2010), but no research has yet examined how reading comprehension relates to successful performance on realistic WPs.

Reported here are a group of studies to investigate elementary school children's ability to take real-world information into account when responding to mathematical WPs. The goal of the studies was to fill gaps in the existing literature by exploring how individual differences – gender, mathematic ability, mathematical reasoning, reading comprehension, and general intelligence – impact performance on WPs. A second goal was to examine the relation between providing realistic responses and success on general mathematical WPs.

The ensuing sections will first provide an overview and discussion of the early research demonstrating the 'suspension of sense making', followed by the attempts that have been made to counter students' lack of sense making. At the same time, although there is a lack of literature investigating individual differences in realistic WPs, individual differences (e.g., gender and general intelligence) will be discussed regarding what role they play in elementary school student's performance on standard WPs.

Suspension of Sense-Making

Wanting to examine the 'suspension of sense making', Verschaffel, De Corte, and Lasure (1994), set out to investigate students' access to, and use of, real-world knowledge during the solution process. To do so, they tested 5th grades students' performance using standard or traditional WPs and problematic or realistic WPs involving all four of the basic arithmetic operations. Standard-items were those asking for a straightforward application of one or more arithmetic operations (e.g., "A boat sails at a speed of 45 km/hr. How long does it take this boat to sail 180 km?" see Table 1, p. 276), while the realistic items are problematic if one seriously considers the realities of the context within the problem (e.g., "John's best time to run 100m is 17 sec. How long will it take to run 1 km?" see Table 1, p. 276). To solve this second item 'correctly', students have to incorporate real-world considerations (i.e., John cannot maintain his top sprinting speed over a whole kilometre, so he must take longer than 170 seconds to run it, although it is hard to know exactly how much longer).

Students were asked to write their numerical answer in the 'answer area' and to provide an explanation as to how they arrived at that answer, or to describe any difficulties they encountered in the 'comments area'. Responses were classified as realistic reactions if either the answer or comment(s) indicated realistic considerations and other answers were considered non-realistic reactions. The students performed poorly on the realistic-items, demonstrating considerable lack of real-world knowledge or consideration during the solution process (also see Greer, 1993). Performance on the standard-items was much better with 84% of these being solved correctly, in comparison to the abysmal performance of just 17% of realistic-items being solved 'correctly' – that is, with the consideration of real-world knowledge (Verschaffel et al., 1994).

Verschaffel et al. also looked at performance on each realistic-item and found considerable item variation. The percentage of realistic reactions was the highest for items that involved a division-with-remainder (DWR) (i.e., a situation where a fractional answer does not make sense). For example: "Grandfather gives his 4 grandchildren a box containing 18 balloons, which they share equally. How many balloons does each grandchild get?" (see Table 1, p. 276). The percentage of realistic reactions (i.e., recognizing that a child cannot have a fraction of a balloon) for this realistic-item was 59%. The authors discuss this higher-than-normal performance as being a result of Flemish students being instructed to answer WPs with a 'response sentence', and by doing so prompted them to consider the appropriateness of their answer.

Similar work using DWR items has been conducted by Silver, Shapiro, and Deutsch (1993), where they used only one WP that varied in whether the computational answer had a remainder equal to one-half, greater than one-half, or less than one-half. They wanted to determine if the remainder size influenced 6th, 7th, and 8th graders' 'sense-making' solutions. Silver and his colleagues found that the remainder size did not have a significant impact on students' ability to make sense of the solution. Overall, only 43% of students responded with the correct, 'realistic' answer, with even fewer (33%) were able to provide a written explanation of their answer as making sense given the context of the problem. This percentage is much higher than that found by Verschaffel and colleagues (1994) overall, but not as high as it was on their DWR item. However, only one question was asked, so the particulars of this question may influence the results. Furthermore, the sample ranged from 6th- to 8th-grade students while Verschaffel et al. studied 5th-grade students. Silver and colleagues argued that students' inability to provide an appropriate interpretation of their solution is due in part to their dissociation of school math performance with reasoning about the real world, coupled with the lack of written explanations required in school mathematics instruction.

Although Verschaffel and colleagues (1994) and Silver and colleagues (1993) provide plausible explanations as to why DWR WPs bring about higher percentages of realistic reactions in children, another possible explanation for the inflated percentage of realistic reactions for DWR problems could be related to personal relevance. Sharing is a form of socialization that is taught to children very early as being necessary to build positive social interactions (Matalliotaki, 2012). Success on these kinds of problems may be primarily due to context relevance (Boaler, 1993, 1994) as young children have plenty of experience with sharing and the difficulties of splitting a balloon.

Caldwell (1995) and Hidalgo (1997) claimed there were very serious methodological limitations in Verschaffel et al.'s (1994) and Greer's (1993) research, and, arguably, the same limitations are found with Silver et al. (1993). First, Caldwell and Hidalgo criticized the use of paper-and-pencil measures to answer the realistic-items, as it does not properly assess what realistic knowledge the children actually possess. Secondly, previous researchers had just assumed that children possessed this real-world knowledge and were not using it, rather than considering students may not possess the real-world knowledge at all.

As a result, subsequent studies have modified Verschaffel and colleagues' (1994) original design, both in the attempt to increase students' use of real-world knowledge or consideration when solving realistic WPs, and to test students' understanding of the necessary real-world knowledge needed to solve the realistic items used in this line of research. For example, De Corte and colleagues (1995) conducted a two-part follow-up study with 5th-grade students. During the first part, 75 students answered the 10 realistic-items, with the purpose of replicating previous findings. For the second part, 64 students answered seven of the original 10 realistic items (see Verschaffel et al., 1994), and the 15 children who responded with the most realistic answers and the 15 who responded with the most non-realistic answers were selected to participate in part two of the study. This second part was an interview, designed both to get a sense of students' awareness and beliefs of the solution process involved in WPs and also to assess the amount of scaffolding needed to convert the non-realistic responders into realistic ones. This modification was necessary to determine if their previous findings (e.g., Verschaffel et al., 1994) were a reflection of students' failure to consider real-world knowledge, lack of real-world knowledge, or simply because they played by the 'rules of the game' and ignored their real-world knowledge.

Students were asked to read aloud a realistic-item followed by their own non-realistic response. The student was then confronted with a confederate classmate's response, which was answered in a realistic fashion. The student was then provided with a weak scaffold (e.g., "What is the best answer? Why?"), followed by a stronger scaffold (e.g., depending on the question, "Draw a diagram of the elements in the problem") in an effort to make the student understand the realistic answer. Part one of the study replicated the results of Verschaffel et al. (1994) with students providing realistic responses only 16% of the time, and of the students who participated in part two of the second study, the top 15 realistic responders only showed 39% and the bottom 15 non-realistic responders showed 8% of realistic reactions. De Corte and colleagues (1995) pointed out that the 15 most realistic responders were not actually performing very well, but they were the "realistic responders" because their responses were more realistic than the rest. After the second, strongest scaffold was given, realistic reactions increased to 57% and this was a cumulative percentage of the realistic and non-realistic problem solvers. The authors did not provide percentages for the change of realistic reactions from the weak to the strong scaffold separately for the realistic and non-realistic responders. This is unfortunate, especially considering De Corte et al. (1995) intentionally isolated the 15 most realistic and non-realistic problem solvers.

Despite De Corte et al.'s (1995) efforts to increase realistic reactions, an improvement to just over half is still not a major improvement. The interview and scaffolding technique was able to show that at least some students do have the understanding of the real-world knowledge or consideration needed to effectively answer the realistic-items (e.g., 16% realistic reactions pre-scaffolding to 57% realistic reactions post-scaffolding) but were resistant to including real-world knowledge into their solution process. De Corte et al. claimed that the culture of school mathematics may be a major contributing factor to low realistic reactions given by students for the realistic-items (i.e., where the culture is for students to just find answers rather than understand the problem). For further explanation see Dewolf, Van Dooren, and Verschaffel (2011).

Wyndhamn and Säljö (1997) were also able to show that children have the necessary real-world knowledge to answer realistic WPs, at least for some problems. When responding to the problem: “Anna and Berra attend the same school. Anna lives 500 metres from the school and Berra 300 metres from the school. How far apart from each other do they live?” (p. 370), the majority of 10- to 12-year-olds responded with ‘it depends’. However, Wyndhamn and Säljö had students complete the task in homogeneous groups of high-, average-, and low-achievers in mathematics formed by their teacher, so it is possible that these students may have responded quite differently than if they had to complete the task independently. Interestingly, the group of high-achievers, and not the other groups, insisted on giving a single answer, even though during their discussion they realized that there could have been more than one answer. This finding very accurately demonstrates the ‘game of word problems’ (Verschaffel & De Corte, 1985, 1997a). The high-achieving students, based on their previous experience, knew that the response ‘it depends’ is not appropriate in the mathematics classroom and that to receive credit for your answer you need to provide one answer. It also further supports the notion put forward by De Corte et al. (1995) that part of the reason why students report so many non-realistic responses to realistic items is due to the culture of the mathematics classroom. The use of realistic thinking, when it comes to math in the math classroom, is not warranted and at times is even frowned upon.

The fact that many children can be guided to give realistic responses suggests that the low percentages of realistic reactions given by elementary and middle school children is not solely due to an absence of real-world knowledge but rather something else. This argument is further supported by the work of Verschaffel et al. (1997) and Inoue (2005, 2008), both showing that educated adults – who can be assumed to have a sufficient degree of real-world knowledge to answer these problems – also report low rates of realistic responses when answering these types of WPs. How can these realistic WPs be designed in such a way as to prompt the use of students' real-world knowledge? The next section will illustrate how researchers have implemented new design features as a way to increase realistic reactions.

Interventions on the 'Suspension of Sense-Making'

In an attempt to replicate and extend upon Verschaffel et al.'s (1994) low percentage of realistic reactions, using a Japanese sample, Yoshida, Verschaffel, and De Corte (1997) had 5th grade students randomly assigned to a neutral or warning condition. The students in the warning condition received written instructions on the top of the test sheet:

> The test contains several problems that are difficult or impossible to solve because of certain unclarities or complexities in the problem statement. When you encounter such a problem, please write it down and explain why you think that you are not able to solve the problem. (p. 333)

All items were literal translations into Japanese of those used in prior research (Verschaffel et al., 1994) and names were changed to reflect those typically given to Japanese children. The realistic reactions of the Japanese students in the neutral condition were not significantly different from Verschaffel et al.'s (1994) original Belgium sample (15% vs. 17%, respectively), nor was there a significant difference in the realistic reactions given between the warning and neutral conditions (15% vs. 20%, respectively). It appears that this type of warning is not sufficient to prompt more students to respond more realistically.

Other warning instructions have been used as a means to increase the amount of realistic reactions provided by students on realistic WPs (Dewolf, Van Dooren, Ev Cimen, & Verschaffel, 2014; Dewolf, Van Dooren, & Verschaffel, 2016; Reusser & Stebler, 1997; Xin et al., 2007; Weyns, Van Dooren, Dewolf, & Verschaffel, 2016). Reusser and Stebler (1997) used a slightly older sample and asked 7^{th} grade students to solve the various realistic-items (Verschaffel et al., 1994) under 4 different conditions: (IC 1) only some standard-items were used, but most were realistic-items answered under the same procedure as Verschaffel et al. (1994); (IC 1A) four of the 10 realistic-items were altered to include a contextual sentence embedded in the problem (e.g., “Think about it carefully before you answer”, or “Study the picture carefully”); (IC 2) all items were accompanied by a set of questions asking for an evaluation of the quality of the problem (e.g., difficulty of understanding and solvability), this condition did not explicitly warn students about the WPs being problematic, it simply directed them towards critically evaluating the problem; and finally, (IC 3) identical to IC 2, plus, in bold print before each set of WPs, students were explicitly told to be cautious: “Be careful! Some of the following problems aren’t as easy as they look. There are, in fact, some problems in the booklet where it is very questionable if they are solvable at all” (p. 319). They found an effect of only IC 1A, that is, 7^{th} grade students provided the highest percentage of realistic reactions on the four realistic-items when a problem-specific warning was used within the WP itself. Given the number of, and variable instruction provided in each condition, Reusser and Stebler’s results indicate that realistic reactions provided by students may not be facilitated by general warning instructions, and rather, warning instructions need to be embedded into each problem for students to begin responding more realistically. Furthermore, it is not enough to include evaluation questions with each WP aimed at cuing students to potential problems (as was done in conditions IC2 and IC3), but rather, explicit warning language needs to be used within each WP to facilitate using real-world knowledge when solving the WP.

This notion is further supported in the work of Xin et al. (2007) who asked fourth-, fifth-, and sixth-grade students to answer Verschaffel et al.'s (1994) realistic-items under two conditions, a 'warning instruction', consisting of a verbal warning that "some of the problems may not be as easy as they seem" (p. 151), and a 'process-oriented instruction', where the students were asked to consider (a) "…[T]he real-life situation behind the problem statement?", and (b) "Is it appropriate to solve these problems by using straightforward arithmetic operations?" (p. 151), both sets of instructions were printed on the front page of each respective booklet. They found a marginally significant difference between the groups (p = .055), with students in the process-oriented instruction condition reporting more realistic considerations (e.g., responses in the comments area), but no differences in the amount of 'correct' answers provided.

Again, the warnings used by Xin et al. (2007) – just as they are in some of the conditions in Reusser and Stebler's (1997) work – are placed at the beginning of the assessment. The findings of these two research studies make it reasonable to think that perhaps more than one prompt is required for students to consider their real-world knowledge as useful when attempting to solve these kinds of realistic problems. However, Dewolf et al. (2014)[1] conducted a very similar study with Belgian and Turkish students using verbal warnings, written warnings at the beginning of the booklet, and

1 Dewolf et al. also used representational illustrations to evoke a real-world scene to which the subject could refer. This manipulation also did not increase realistic reactions.

additional shortened written versions of the warning in the header section of each subsequent page, and still found that students' percentage of realistic reactions on Verschaffel et al.'s (1994) realistic-items were no better than students being given no warning.

These results were not what the authors had anticipated. Surely, if you at least inform students of the possibility of encountering a problem, they could implement a different solution strategy to approach the problem or consider the adequacy of the given answers. But, when you explicitly tell them that, "There are, in fact, some problems in the booklet where it is very questionable if they are solvable at all" (Reusser & Stebler, 1997, p. 319), it is quite surprising to see that they do not heed the warning, and instead continue to respond using conventional methods. There must be another way to alter the problem or the context of the experimental setting to increase the realistic reactions given by students. The preceding studies have indicated that, to elicit realistic reactions in students, it requires more than simply telling them to pay attention to the problem or warning them that there could be potential problems.

DeFranco and Curcio (1993) were able to increase the number of realistic reactions given by 6th grade students through enhancing the *authenticity* of the problem and situation. Enhancing the authenticity is accomplished by altering the problem to better simulate some of the aspects of the real-life task situations by providing more descriptive information about the task context and the purpose of solving the problem (Palm, 2008; Palm & Nyström, 2009; also see Zwaan & Radvanksy, 1998 for similar information on Situation Models). In part one of their study, they asked students to review an already answered WP (e.g., "328 senior citizens are going on a trip. A bus can seat 40 people. How many buses are needed so that all the senior citizens can go on the trip", p. 61), the answer provided was 8r8 or 8.5. The task was stated as open-ended. This allowed students to critique the work of other classmates and discuss whether they agreed or disagreed with their peers' solution. During the individual interview, students were asked if they agreed with the answer given or if they would give a different answer. Only 10% of students (N = 20), when asked if the original answer of 8r8 or 8.5 was correct or if they believed another answer was correct, responded with changing the original answer to an answer of 9 instead (the realistic response). The other 90% of students erred with keeping the original answer, rounding down to 8, or making a calculation error. In part two of the study – one month later – the same students were given this information:

> FACTS:
> *Date of party*: Friday, April 15
> *Time*: 4:00 – 6:00 PM
> *Place*: Ricardo's Restaurant, Queens
> *Number of children attending the party*: 32
> PROBLEM:
> We need to transport the 32 children to the restaurant so we need transportation. We have to order minivans. Board of Education minivans seat 5 children. These minivans have 5 seats with seatbelts and are prohibited by law to seat more than 5 children. How many minivans do we need?
> Once you have determined how many minivans we need, call 998-2323 to place the order (p. 62).

The students were asked to complete the problem and make a telephone call to place the order. At that point, the researcher left the room to 'see another student'; the researcher actually left the room to be on the other end of the telephone to complete the student's order.

When responding to this problem, student's realistic reactions increased to 80%. Although numerically different, this WP, is essentially the same type of problem (i.e., a division-with-remainder problem). It is possible that the increase of realistic reactions is an artifact of part one, such that students remembered discussions with their peers regarding a similar problem. This explanation is unlikely, as the students were interviewed after the opportunity to discuss the problem with their peers in part one of the study and 90% still reported an unrealistic answer. It is possible that the increase of realistic reactions in part two is due to possible feedback provided post-interview. If this is the case, any feedback provided to the students was not documented in the original publication. Overall the change from part one to part two showed that 80% of students were able to solve the problem correctly when the authenticity of the situational context was enhanced – the 10% who were correct in part one were also correct in part two, and 20% of the students were incorrect in both parts.

Part one of the study was very typical of a mathematics classroom, which is why it was called the restrictive context. The students' main concern in part one was with the computation needed to solve the problem. While in part two of the study, students had to physically make a phone call to complete the problem. It is this feature coupled with the information in the problem (e.g., 32 students were in this particular 6^{th} grade class) that made this context more authentic. The number of realistic reactions significantly increased from the restrictive to the authentic setting, and it was thought by the authors that this occurred because the students had to reconsider the fractional answer in light of not being able to order a fraction of a minivan (DeFranco & Curcio, 1993).

Students seem to ignore the warnings provided with realistic WPs and only when the context is fundamentally changed do students show an inkling of using realistic knowledge and consideration when solving these kinds of WPs. It could be that students simply do not have the skills necessary to answer such problems. For example, mathematical curriculum textbooks rarely if ever include WPs that have more than one answer or no answer, WPs with extraneous information, WPs that are missing information, or even WPs that involve more than one step of calculations (Verschaffel et al., 2000). Given the context that students learn in, it becomes less surprising that students are reporting extremely low percentages of realistic reactions to WPs that require real-world knowledge. The typical mathematics classroom and teacher do not facilitate common sense in math (see Chacko, 2007; Verschaffel et al., 1997), WPs are often perceived as artificial, puzzle-like tasks with no actual connection to the real-world (Verschaffel & De Corte, 1997b; Verschaffel et al., 1994), and in many cases providing realistic consideration can be more harmful than helpful in arriving at the 'correct' answer (Verschaffel et al., 1997, Xin et al., 2007).

Elementary school children are resistant towards including their realistic knowledge into the solution process of a mathematical WP, and even when explicit warnings are given, students are still reluctant. The verbal and written warnings do not seem to be enough to provide a substantial increase in realistic responding. What about pictorial representations accompanied with WPs? If students are given a visual representation of the problem situation are they better able to incorporate their real-world knowledge? In a series of experiments (Dewolf et al., 2014; Dewolf, Van Dooren, Hermens, & Verschaffel, 2015; Dewolf et al., 2016; Weyns et al., 2016) with 5^{th} and 6^{th} grade students, researchers have attempted to increase the number of realistic responses to WPs that require the use of realistic considerations by including pictorial representations of the situations described in the WP. In some cases, students were given a warning paired with an illustration (Dewolf et al., 2014), and in other cases the realistic elements of the pictures were included and or highlighted. For example, in the planks problem "Steve has bought 4 planks of 2.5m each. How many planks of 1m can he get out of these planks?" (Verschaffel et al., 1994, p. 276) Dewolf and colleagues (2016) provided an illustration that showed a basket for waste pieces, and or the waste basket was highlighted.

When the images were paired with a pictorial representation of the situation 5th grade students did not perform any better on the realistic items than if they had not received a picture (Dewolf et al., 2014). Even when cueing elements were used (e.g., highlighting specific actions) to guide student's attention towards the pieces of information to assist with the problem solution, 5th and 6th grade student's realistic responses did not improve (Dewolf et al., 2016; Weyns et al., 2016). When Dewolf and colleagues (2015) sought to determine which elements students were focusing on in the pictures, they found that students were not even looking at the pictures, and when the pictures were presented in a way that forced students to view them (e.g., WPs and illustrations were presented sequentially rather than simultaneously), the pictures still did not improve their realistic responding.

The major findings in this line of literature are that children rarely consider real-world implications when solving mathematical WPs that require real-world consideration. Furthermore, this finding is not because students do not have access to the real-world knowledge needed to properly approach these kinds of WPs. Various methodological techniques have been created in the hopes to increase the amount of realistic responses provided by students to these realistic WPs, however, most have shown to be ineffective in creating substantial improvements. Despite the low percentage of realistic reactions provided by students, there are those who do use their real-world knowledge. Who are these students, what makes them different from the students who do not use their real-world knowledge, and is it beneficial in their academic pursuits? These questions will be addressed in the next section on individual differences.

Individual Differences

Verschaffel et al. (2000) point out that "a limitation of the research is that…relatively little attention has been paid to individual differences" (p. 156). Verschaffel et al. are referring to both their own research and to this particular area of research in general. However, there have been both direct and indirect contributions made to individual differences. The following section elaborates on research investigating potential individual differences such as reading comprehension (Bjork & Bowyer-Crane, 2013; Muth, 1984; Pimperton & Nation, 2010), gender (Cooper & Dunne, 2000; Boaler, 1993, 1994; Palm & Nyström, 2009), and prior mathematical ability such as mathematical achievement and calculation ability (Bjork & Bowyer-Crane, 2013; Kail & Hall, 1999) that play a role in children's ability to successfully solve realistic and standard WPs.

One of the points of focus in this paper is how reading comprehension is linked to performance in WPs that require the consideration of real-world knowledge to effectively answer the problem. As mentioned above, DeFranco and Curcio (1993) altered the authenticity of the context (e.g., providing detailed information and personal relevance with the number of students attending the event) and the authenticity of the situation (e.g., having students actually make the phone call to order the transportation). One of the pitfalls of DeFranco and Curcio's design is they made no attempt to distinguish the difference between the context and the situation and this could have been achieved simply by having a group of students receive only the enhanced authenticity of the context, (i.e., students who would have seen the information but did not have to physically make a phone call). Altering the authenticity of one aspect – the context of the problem – can improve students' performance on realistic WPs (see Palm, 2008; Palm & Nyström, 2009). Creating a more authentic context or enhancing the situation model (see Zwaan & Radvanksy, 1998), means students are better able to represent the text as a mental model, which essentially 'paints a picture' of the problem in their minds. This brings us to the next topic of discussion, reading comprehension.

Reading Comprehension

Previous research has reliably shown an association between reading ability and standard mathematical WP solving performance in children (Bjork & Bowyer-Crane, 2013; Muth, 1984; Pimperton & Nation, 2010; Vilenius-Tuohimaa, Aunola, & Nurmi, 2008), but this line of research has not specifically addressed realistic WP solving. Reading comprehension, as it relates to general mathematical WP solving performance, has been examined in elementary and middle-school children (Bjork & Bowyer-Crane, 2013; Pape, 2004; Vilenius-Tuohimaa et al., 2008), adolescents (Kyttälä, & Björn, 2014), with children of average and poor reading comprehension (Pimperton & Nation, 2010), and how reading comprehension, via working memory, impacts WP solving performance (Lee, Ng, Ng, & Lim, 2004; Swanson, 2004). Because of the strong relation between reading comprehension and WP solving ability in general, it is worth reviewing these findings, as they may also relate to realistic problem solving.

Bjork and Bowyer-Crane (2013) compared the ability of 6- and 7-year-old children to solve WPs to their ability to solve numerical operations, and also examined if these differences were related to different cognitive skills. Students were tested on phonological awareness (e.g., identifying and manipulating units of verbal language), verbal ability, reading accuracy, and reading comprehension. They were also given 15 numerical operation problems that were matched to 15 WPs. The children in this study performed significantly better on the numerical operations task than on the WPs task. The authors suggest this finding indicates that the WPs task demanded more cognitive skill than the numerical operations task. This led Bjork and Bowyer-Crane to use regression analyses to determine which of these cognitive skills were important predictors for both types of mathematical skills. The results of these regression analyses determined that only phonological awareness was a significant predictor of numerical operations performance. However, phonological awareness, reading comprehension, and numerical operations were all significant unique predictors of performance on the WPs task, after verbal ability (i.e., expressive vocabulary – providing definitions to progressively more difficult words – as measured by the Wechsler Abbreviated Scale of Intelligence) was already accounted for. This suggests that reading comprehension is an additional cognitive skill needed for solving word problems.

Utilizing a similar-aged sample of students as Bjork and Bowyer-Crane (2013), Pimperton and Nation (2010) went a step further by categorizing the students based on their relative reading ability. Reading ability was assessed using both reading comprehension and reading accuracy, and students were divided between Poor Comprehenders and Control. Poor comprehenders were classified as those students who scored low in reading comprehension but were average or above average in reading accuracy skills. Students in the control group were matched with students in the poor comprehenders group on non-verbal ability, multiple measures of reading accuracy, and chronological age. Furthermore, the students in the control group had scores on the reading comprehension measure that were equal to or greater than their scores on the reading accuracy measure. Mathematical ability was assessed using the Wechsler Individual Achievement Test–II, which included the subtests mathematical reasoning and numerical operations. Mathematical reasoning was assessed through a series of verbally and visually presented problems such as creating and solving computation problems, patterning problems, using graphs and grids to make comparisons, answering one-step and multi-step word problems (WPs were presented orally, however, they remained visible in front of the student so as to not tax their working memory), etc. Numerical operations was assessed using a paper-and-pencil test measuring untimed written math calculation skills, where students were required to solve written calculation problems that increased in difficulty. This was said to assess procedural fluency.

Using a mixed design ANOVA with numerical operations versus mathematical reasoning as the within factor, Pimperton and Nation (2010) found a significant interaction. The poor comprehenders scored significantly worse than the control group on the measure of mathematical reasoning but scored no differently on the numerical operations task. Regression analyses showed that once expressive vocabulary (i.e., definitional knowledge of increasingly difficult words) was accounted for in a hierarchical model, group membership (poor comprehenders vs. control) did not account for any additional variance. Pimperton and Nation (2010) interpreted this result to mean that poor comprehenders' challenges in math are more a consequence of language problems, rather than mathematical reasoning, and that language ability is necessary to reason mathematically.

Vilenius-Tuohimaa et al. (2008) also examined the relation between reading comprehension and math WP skills; however, their sample of fourth graders (9- to 10-year-olds) was a bit older than the samples used by Bjork and Bowyer-Crane (2013) and Pimperton and Nation (2010). Children were categorized as good or poor readers based on their technical reading skill, which was assessed using the word recognition subtest of a norm-referenced Finnish reading test known as the Ala-Asteen Lukutesti (ALLU) Reading Test for Primary School. Students had to separate words written in sets by marking a line between each word (e.g., 'racetowardspairunder' would be 'race/towards/pair/under') – this assessed speech and accuracy. Poor readers were defined as those who scored 1-2 *SD*s below the mean. In Vilenius-Tuohimaa et al.'s sample, 29.8% of students were categorized as poor readers and the remaining 70.2% categorized as good readers. For reading comprehension, students read four different texts followed by 12 multiple-choice style questions. Mathematical WPs were taken from the NMART Counting Skills Test. The students had to answer 20 problems and received one point for each correct answer.

Not surprisingly, children in the good reading group performed significantly better than children in the poor reading group on the reading comprehension measure, but those in the good reading group also scored better on the math WPs task. By means of path analyses, technical reading score was found to be a significant predictor of math WP solving performance accounting for 18% of the variance in WP scores. Technical reading score was also a significant predictor of reading comprehension, accounting for 23% of the variance in reading comprehension scores. When technical reading scores were controlled for, reading comprehension was still significantly related to math WP solving. The authors suggest this finding indicates that both of these skills – reading comprehension and math WP solving performance – involve an overall reasoning component (Gelman & Greeno, 1989, as cited in Vilenius-Tuohimaa et al., 2008).

Kyttälä and Björn (2014) examined the relation between literacy and math WP performance in adolescents, using a sample of 8th grade students, who had to answer questions on numerical calculations and solve math WPs. Students were also assessed on reading comprehension and technical reading skills, using subtests of the Dyslexia screening test for adolescents and adults. It was found that literacy skills were the best predictors of WP performance, however, there was an interaction effect found with gender. Reading comprehension skills only predicted WP performance for boys. Girls had higher levels of both reading comprehension and technical reading skills than boys in this sample. However, as this study was with adolescents and most of the previous work has focused on elementary school children, it is an open question if this same gender effect would be evident in a younger age group.

The above-mentioned studies were all aimed at examining how reading comprehension is related to performance in mathematical WPs in general and have not examined reading comprehension in regard to realistic WPs. There is some research, however, that is explicitly about realistic WP that can also be considered to be indirectly about reading comprehension. This research aims to change the written description of WPs to increase their "authenticity".

Authenticity

Palm (2008; Palm & Nyström 2009) expanded on Verschaffel et al.'s (1994) original work by using only their realistic items, and modified them to be more authentic; that is, the more authentic problems were considered to better simulate some of the aspects of real life task situations by providing more descriptive information about the task context and the purpose of solving the problem. As such, authenticity is one of the few individual differences that has been explicitly tested with realistic WPs. For example, here is the plank problem used by Verschaffel et al. (1994): "Steve has bought 4 planks of 2.5m each. How many planks of 1m can he get out of these planks?" (p. 276), while the more authentic version of this problem, used by Palm (2008) is:

> You are building a cabin and as walls you want to use planks that are 1m long. You are at the moment short of thirteen 1-meter planks. A friend says that she has found 4 planks, each 2.5m long. You are wondering if that is enough to finish the walls. How many 1-meter planks can you saw out of the planks she found? (p. 66)

As can be seen from the example of an authentic problem, there is more descriptive information, such as the purpose of needing additional planks and why the planks are required to be 1m.

Students' responses were categorized as realistic if their answer followed a solution process that involved the appropriate use of knowledge about the real-world situation. Non-realistic answers were all those that did not fit the criteria for realistic answer. Students were also interviewed afterwards and their responses in the interview were coded as providing realistic considerations, independent of whether their answer reflected these considerations. All other information was coded non-realistic consideration. Palm (2008) found an increase in performance (51% of realistic responses) of 5th graders on realistic WPs by enhancing the authenticity of the problem. Surprisingly, Palm found a much higher percentage of realistic responses to the less authentic (e.g., the realistic WPs found in the majority of the literature) WPs in his sample in comparison to the bulk of the literature (e.g., 33% vs. 16-17%, respectively).

The way Palm explains enhanced authenticity shares similarities to the dimensions of what Zwaan and Radvanksy (1998) call a *situation model* (SM), that is, an integrated mental representation of described states of relationships that are amalgamated from information stated explicitly in the text and information already known. A situation model specifies that any situation described through text can be characterized by five dimensions (protagonist, causation, motivational, temporal, and spatial, see Zwaan & Radvanksy, 1998 for detailed descriptions of the five dimensions). Although SMs are used to help explain general text comprehension (Zwaan & Radvanksy, 1998), they have also been applied to the understanding of math WPs (Coquin-Viennot & Moreau, 2007; Vicente, Orrantia, & Verschaffel, 2008; Thevenot, Devidal, Barrouillet, & Fayol, 2007).

In light of Zwaan and Radvanksy's (1998) dimensions of situation models and the work done in the field of mathematics using situation models as a theoretical framework, it can be argued that Palm's (2008; Palm & Nyström, 2009) manipulation of authenticity, although not an explicit attempt to alter reading comprehension, may have inadvertently created a WP that elicits a richer SM and thereby enhances students' ability to comprehend the text. Palm suggests the more authentic versions of the realistic WPs allowed the students to be more engaged in the task context and resulted in more willingness to disregard the rules of the classroom. An alternative explanation could be that the more authentic versions of the realistic WPs enhanced students' comprehension of the text resulting in an improved mental representation of the situation, and it is their improved comprehension that encouraged them to consider the relevance of the realistic nuances described in the WP.

Mathematical Ability

Arithmetic knowledge is widely agreed upon as being essential for solving mathematical WPs (Kail & Hall, 1999; Muth, 1992; 1984; Zheng et al., 2011). *Arithmetic* is the addition, subtraction, multiplication, or division of single-digit numbers or in their simple forms (e.g., 2 + 3 = 5, 25 ÷ 5 = 5, Bjork & Bowyer-Crane, 2013; Fuchs et al., 2006). Students feel more comfortable with and perform better when solving arithmetic calculation problems in comparison to arithmetic WPs (Muth, 1984). For example, in the Bjork and Bowyer-Crane (2013) study mentioned above, 6- and 7-year-olds completed 15 arithmetic calculations and 15 WPs that had the exact same numbers and operations as were used in the arithmetic calculations (i.e., the calculation problem '9 + 5 =' corresponded to the matched WP, 'Ann has 9 pennies and her friend has 5 pennies. If they add their pennies together, how many do they have?' p. 1351). The students performed significantly better on the arithmetic calculation problems compared to the arithmetic WPs.

In a sample of 6^{th} graders, Muth (1984) found that computation ability was strongly and positively correlated (i.e., $r_{(198)} = .60$) with the number of correct answers on the 15 arithmetic WPs (Muth, 1984). Through a regression analysis, Muth also found that, after the contribution of reading ability was included, computation ability accounted for an additional 7.62% of the variance in correct answers on the WPs task. Fuchs et al. (2006) found results similar to those found by Muth (1984) using a sample of 3^{rd} graders, although the correlation coefficient was somewhat smaller (i.e., $r_{(310)} = .45$). Fuchs et al.

used path analyses to determine that arithmetic ability was a significant independent predictor of arithmetic WP solving performance. Furthermore, arithmetic skill remained a significant independent predictor of arithmetic WP performance after phonological decoding and sight word efficiency were removed from the model, and again when language was removed from the model. These results strengthen the argument that procedural fluency in addition and subtraction is a foundational skill for subsequent mathematical skills, such as solving arithmetic WPs.

In a number of other studies, arithmetic calculations have been shown to be predictive of performance in standard mathematical WPs for students spanning the elementary grades (e.g., 2nd – 6th grade, Bjork & Bowyer-Crane, 2013; Kail & Hall, 1999; Swanson, 2004; Zheng et al., 2011). Zheng et al. (2011) examined arithmetic calculation as a mediating factor between components of working memory (WM) and WP solving performance. Examining students in 2nd, 3rd, and 4th grade, Zheng and colleagues assessed arithmetic calculation using a subtest of the Wechsler Individual Achievement Test (WIAT) and the Wide Range Achievement Test-third edition (WRAT-III), where the number of problems correct was the students score. The WP solving task was assessed by having students provide a mental solution to orally presented WPs from the Wechsler Intelligence Scale for Children-third edition (WISC-III). The problems were given to students orally because reading ability was considered to confound the contribution of working memory to problem solving. The mediation analysis indicated that math calculation significantly mediated the relation between the central executive component (a facet of working memory) and solution accuracy in WPs (standardized indirect coefficient = .19, *SE* = .09). The authors further state that this mediation fully accounted for the relation between the working memory components and solution accuracy. Swanson (2004) also looked at working memory as a predictor of children's mathematical WP solving performance. He found calculation skills to be predictive of WP solving performance, and once working memory was controlled for, calculation skills contributed a significant unique amount of variance (16%, $p < .05$).

Thus far, the preceding section has demonstrated that calculation ability is positively correlated with WP solving performance (Bjork & Bowyer-Crane, 2013; Kail & Hall, 1999; Muth, 1984; Zheng et al., 2011), calculation ability is a predictive factor for WP solving success (Fuchs et al., 2006; Swanson, 2004), and that calculation ability is a mediating factor in the relation between working memory components and WP solving performance (Zheng et al., 2011). All of these studies examined WP solving performance using standard mathematical WPs. This paper however, focuses on children's ability to solve realistic WPs. One study, however, has examined how general math ability is related to both standard and realistic word problem solving. In a sample of 4^{th} and 6^{th} grade Chinese students, Xin and Zhang (2009) administered a WP task with both standard and realistic WPs. The realistic WPs were modified from Verschaffel et al.'s (1994) problems such that the structure and content of the WPs were held constant, but they were translated and names were changed to be common Chinese names. Mathematical achievement data came from the previous semester's final exam and measured both procedural and conceptual mathematical knowledge at each child's grade level. Procedural knowledge refers to the 'how', the child's ability to perform calculations to arithmetic problems (e.g., solve 3x – 4 =). Whereas conceptual knowledge refers to the 'why', the child's ability to understand concepts (e.g., which is bigger 5x or

8x). The results demonstrated that prior mathematical achievement was a significant predictor of success on both standard and realistic mathematical WPs.

Gender

The National Mathematics Advisory Panel (U.S. Department of Education, 2008) reports that gender differences are small to nonexistent, and rather it is society's fixation on gender differences that continue to drive this area of research, rather than attempting to raise the mathematical scores of both boys and girls. Gender differences that have been found tend to be in specific areas within the domain of mathematics rather than the whole of mathematics (Delgado & Prieto, 2004; Friedman, 1995; Hyde, Fennema & Lamon, 1990). Spatial reasoning (e.g., mental rotation) shows the largest and most consistent difference between boys and girls, favouring boys (Boonen, van der Schoot, van Wesel, de Vries, & Jolles, 2013; Delgado & Prieto, 2004; Friedman, 1995; Johnson, 1984), boys also have a tendency to outperform girls as they get older (Friedman, 1995; Hyde et al., 1990), however there is more variability among boys, such that there are more boys found in both the lower and higher scoring range of many assessments (Delgado & Prieto, 2004). In regards to WPs, both high school (Hyde et al., 1990) and elementary school (Lummis & Stevenson, 1990) boys perform slightly better than girls. It remains unclear as to whether this is a genuine difference or a result of mathematical beliefs and gender stereotypes. Children and parents alike have a misguided belief – which can be seen as early as 1st grade – that boys are better than girls at math (Cvencek, Meltzoff, & Greenwald, 2011; Lummis & Stevenson, 1990), and boys as young as six years-of-age also identify more strongly with math than girls on implicit and self-report measures (Cvencek et al., 2011).

Boaler (1993, 1994) showed that the context (i.e., a gender bias) of a WP could have differing impacts on boys and girls, more specifically that stereotypically female contexts can negatively affect girls (Boaler, 1993, 1994). Eighth-grade students were asked to complete WPs with a stereotypical gendered context (e.g., a wood cutting problem, a fashion workshop problem, or a football related problem), as well as those with a gender-neutral context and those that were abstract problems that provided no context. Boaler also tested two different schools – one where instruction separated content and process (e.g., concentrating upon content in school and process at home), treated homework as something only done at home and not talked about in school, and discouraged students from making real-world assumptions when solving WPs, and another school that that worked on open-ended activities in a discussion-oriented environment. While there were no differences between boys and girls in the second school, there were differences between boys and girls in the first school. Most boys drew a diagram to represent the wood cutting problem (Boaler, 1993, 1994), whereas all of the girls in the first school only used numbers and words. Interestingly, for a problem embedded in the context of a fashion workshop, girls performed worse than boys, but, again, only in the first school that disconnected content and process. In this problem, students were given a list of tasks that needed to be done and the number of hours needed to do them and were asked to optimally assign the tasks in order to get the most work done. Girls were more likely to take into account that certain tasks could not be done until others were completed (e.g., cannot make deliveries until the clothes have been made) and, for that reason, could not get all the jobs done. Boaler suggests that these girls had a difficult time disengaging from the context and, therefore, responded to the problem using their real-world knowledge and reported an “incorrect” answer. The second school approached mathematics in a more open-ended fashion, allowing and encouraging students to explore and discover the mathematical problems. There were no set rules on using particular processes or methods, so students were able to employ any route possible to find the appropriate answer. The girls in this school were not as negatively impacted by the real-world context (i.e., like that of the fashion workshop). Boaler (1994) suggests the open communication and negotiation of mathematical processes valued at that school allowed a less threatening environment combating the underachievement typically found in the performance of girls in mathematics.

The realistic answers given by the girls in the school that disconnected content and process were considered incorrect because the realistic WPs in this study were not the typical realistic-items that have been seen in the majority of the literature (e.g., Greer, 1993; Verschaffel et al., 1994), but rather they were taken from National Assessment Tests. The fashion item on this test, despite having the feel of a problem that required realistic consideration, was not actually designed to require real-world consideration, and it is simply a WP that required the test taker to find the optimal number of combinations that resulted in the smallest amount of time. Perhaps students are supposed to assume that the deliveries could be of materials that were not yet made or were simply not supposed to think of such things. The fashion workshop problem is an abstract mathematical problem blanketed in a real-world context. The girls in the sample incorrectly used their real-world knowledge (e.g., having patterns cut and sewn before having them shipped) to answer this problem, however, the problem was based on an abstract mathematical concept, which ignored the order at which jobs should be performed. What this indicates is that the WPs – like the fashion workshop problem – which were relevant to girls, produced realistic thinking when it was not warranted. This was not the case for boys when faced with the wood cutting problem. Instead, this problem did require the use of real-world knowledge, and the boys were better equipped with this knowledge in comparison to the girls, allowing them to see that it was more appropriate to provide an answer that considered a full length 1m plank and not two 0.5m planks as being equivalent to a 1m plank. Boaler cites previous research by the Assessment of Performance Unit (Murphy, 1990; as cited in Boaler, 1993, 1994) suggesting that girls may have more difficulty than boys in abstracting issues from their context, because girls value the circumstances that a task is presented in.

When examining WPs that have been designed to require real-world knowledge and consideration, gender differences are not as clear. As mentioned above it the section on authenticity, Palm and Nyström (2009) expanded on Verschaffel et al.'s (1994) original work by using only their realistic-items, and modified them to be more authentic, that is, the more authentic problems were considered to better simulate some of the aspects of real life task situations by providing more descriptive information about the task context and the purpose of solving the problem. There were no differences between boys and girls when activating real-world knowledge for both the less authentic and more authentic types of WPs. Both boys and girls did show significantly greater proportions of realistic answers and considerations on the more authentic WPs in comparison to the less authentic version of the WPs [Verschaffel et al.'s (1994) realistic-items]. For example, boys' realistic answers and considerations increased from 31% to 46% and 42% to 50%, respectively; girls' realistic answers and considerations increased from 34% to 53% and

42% to 57%, respectively. Palm and Nyström (2009) only found gender differences for one particular item:

> You are going on a camp for 4 days, but you also want to ride. Your dad sees in the camp papers that you have 45 minutes free time each day, and that horses can be rent for tours round a path in the woods that takes 10 minutes. To know how much money you shall bring you must calculate how many tours you have time to ride. How many 10 minutes tours do you have the time to do during these days? (p. 66).

Girls benefited more than boys on this particular item from the higher task authenticity. These findings both support and conflict with those found by Boaler (1993, 1994). Contrary to Boaler's results, Palm and Nyström (2009) did not find significant evidence to suggest that girls were more likely overall to be affected by the context of the problem and thereby show greater activation of real-world knowledge and consideration. However, Palm and Nyström (2009) did find that the gendered context of some problems did impact the girls differently than the boys across the standard and authentic WPs, such that the girls gave more realistic knowledge and consideration for problems that were contextually relevant to their lives or interests (e.g., the horse riding problem) and vice versa, the girls were less likely than the boys to use real-world knowledge and consideration for gendered contexts not relevant to their lives or interests (e.g., plank problem). The girls gave three times as many realistic answers for the more authentic horse riding problem than the less authentic version, but not even twice as many realistic answers for the more authentic plank problem than the less authentic plank problem. The boys however, provided twice as many realistic answers to the more authentic versions of both the planks and horse riding WPs in comparison to the less authentic versions. So, depending on the context and the WP, the girls in Palm and Nyström's study were differentially impacted by the context of the problem in terms of the amount of realistic consideration they gave to solving the problem and the answer they provided.

It is important to note that three of the 10 less authentic WPs used in Palm and Nyströms study were explicitly gendered, that is, gendered names were used in the problem and that problem asked what the gendered character would do (e.g., Anton needs planks of 1 m, Elin rides horses, how many 10 minute rides does *she* have time for, and Martin's best time to run 100 m is 10 sec). When it came to the more authentic WPs, all but one eliminated any explicit reference to gender, specifically the running problem:

> There is an athletics competition on TV. You and a friend watch when the fastest man in the world, Maurice Green, wins the 100 m race on the time 10.00 sec. The next race you watch is 10000 m, which is won by Haile Gebrselassie on the time 26 min. and 5 sec. What do you answer when your friend asks you: What time do you think it would take Maurice Green to run 10000 meters (= 1 Swedish mile)? (p. 66).

This problem is gendered because the names used in the problem are the names of men. The remaining more authentic WPs addressed the reader with, 'you' as opposed to 'she', 'he', 'him', 'her', etc. Palm and Nyström's explanations for their findings of gendered context, as seen above, are based solely on whether they believe the information in the problem as relating to a boy or girls interest. The less authentic items used by Palm and Nyström have the potential confound of gendered information. The lower percentage of realistic answers and considerations for those particular items could be due to the reduced authenticity, or the gendered context. The authors, however, believe the poor results are due to the lack of authenticity.

Csíkos (2003, as cited in Verschaffel, Van Dooren, Greer, & Mukhopadhyay, 2010), found no difference between boys and girls in the amount of realistic responses given, except on the planks problem, where boys produced a significantly greater percentage of realistic responses than girls (19% vs. 8%, respectively). Csíkos provides the same explanation; boys might have more real life experience or interest in sawing planks. Cooper and Dunne (2000) have found evidence supporting Boaler's notions, specifically that girls are performing less well on National Curriculum Assessment items using realistic problems, for similar reasons. This is in direct contrast to gender results reported by Dewolf et al. (2014), who found no gender differences in response to realistic problems overall, no gender differences when given a warning about problematic items, and no gender differences when providing an illustration to help with reasoning when solving realistic WPs. And finally, as an interesting note, Cooper and Harries (2002) found that the boys in their sample were slightly, but not significantly, more likely than girls (17.9% vs. 11.1%) to show a willingness to 'break the rules' when solving a DWR problem. The original WP states that the maximum capacity for an elevator is 14 people, but the boys in this sample would more often allow more than 14 people onto the elevator in their attempts to determine how many trips the lift must make to transfer 269 people from the bottom floor to their respective floors during the morning rush hour.

Currently there is no consensus on whether boys and girls differ in their ability to use their realistic knowledge base when solving WPs that require realistic considerations. The research examining gender differences is inconsistent at best. What does seem to be a consistent finding is that the stereotypical gendered context does play a role in girls' and boys' performance on realistic WPs (Boaler,1993, 1994; Csíkos, 2003 as cited in Verschaffel et al., 2010; Palm & Nyström, 2009). In some cases, it appears that the stereotypical gendered context is a greater hindrance for girls (Boaler, 1993, 1994; Cooper & Dunne, 2000). However, Boaler (1993, 1994) has also shown that these gender differences are less prominent in educational institutions that focus on communication through open-ended activities in a discussion-oriented environment. This environment is said to be less threatening, especially for girls (Boaler, 1993, 1994).

As mentioned at the beginning of the section on gender differences, continuing to examine this relation may hurt students in the long run. Therefore, the worst-case scenario is that gender differences are an irrelevant factor in students' use of realistic reasoning when solving math word problems. It is especially important to consider that some of the major gender differences presented in the literature (e.g., Boaler, 1993, 1994) are found in word problems that depict a real-world situation, but still do not require the student to use their realistic reasoning. There remain many questions to be examined when it comes to differences between boys' and girls' use of realistic reasoning.

The Current Studies

This dissertation consists of three studies reported in two chapters. The first chapter describes the first study, and it focuses on examining individual differences that may play a part in elementary school children's performance on word problems that require the use of one's real-world knowledge and consideration. With the extensive research to date on the relation between reading ability and mathematical WP performance, reading comprehension is an obvious point of interest for children's realistic WP solving performance. It also seems reasonable, given the existing body of literature, that mathematical ability is an important component of successful realistic WP performance in general, or that different aspects of mathematical ability (e.g., math reasoning vs. calculation ability) might be differently related to success on realistic WPs. On the other hand, it is also possible that, for realistic WPs, that math ability will no longer be related to performance after controlling for verbal ability. In addition, the research on gender differences is suggestive but not conclusive. To that end, the first chapter examines factors such as gender, non-verbal intelligence, reading comprehension, mathematical ability, and math reasoning, assessed using the Raven's Standard Progressive Matrices and the Woodcock-Johnson III Tests of Achievement (WJ III Ach). This chapter will also examine if using one's real-world knowledge is beneficial to general word problem solving.

The second chapter presents three intervention studies aimed at increasing the number of realistic responses given by students on realistic word problems. In the case of Study 1, these are the same participants presented in the first chapter, but this chapter focuses on the results of the intervention that was part of this study while the previous chapter only considered individual differences. Study 1 manipulated the way students provide their numerical responses, with some reporting a numerical answer and others are asked to use a response sentence. In Study 2, the instruction are manipulated – some students receive the basic instructions and others receive the instructions with examples. Lastly, Study 3 attempts to increase realistic responses by enhancing the authenticity of the problems (see Palm, 2009) and pairs the word problems with multiple-choice reading comprehension-like questions as a way to ensure students are actually reading the entire text.

This is manuscript-style dissertation. This means that each data-reporting chapter (i.e., Chapters 2 and 3) is a stand-alone piece that is able to be submitted to a journal as its own manuscript. The implications of this is that each chapter will have its own reference list, and also there will be some repetition of content across the chapters. The last chapter will provide a general discussion of the implications of all of the studies, and will discuss some additional analyses (included in Appendix G) not reported in Chapters 2 and 3.

References

Bjork, I. A., & Bowyer-Crane, C. (2013). Cognitive skills used to solve mathematical word problems and numerical operations: A study of 6- to 7-year-old children. *European Journal of Psychological Education, 28*, 1345-1360. doi:10.1007/s10212-012-0169-7

Boaler, J. (1993). Encouraging the transfer of ‘school’ mathematics to the ‘real-world’ through the integration of process and context, context and culture. *Educational Studies in Mathematics, 25*, 341-373.

Boaler, J. (1994). When do girls prefer football to fashion? An analysis of female underachievement in relation to ‘realistic’ mathematic contexts. *British Educational Research Journal, 20*, 551-564.

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*, 271-279. doi:10.1016/j.cedpsych.2013.05.001

Caldwell, L. (1995). *Contextual consderations in the solutions of children’s multiplication and division word problems.* Unpublished undergraduate thesis, Queen’s University, Belfast, Northern Ireland.

Chacko, I. (2007). Real-world problems: Teachers’ evaluation of pupils’ solution. *Studies in Educational Evaluation, 33*, 338-354. doi:10.1016/j.stueduc.2007.07.008

Cooper, B., & Dunne, M. (2000). *Assessing children’s mathematical knowledge: Social class, sex and problem-solving*. Buckingham, Philadelphia: Open University Press.

Cooper, B., & Harries, T. (2002). Children's responses to contrasting 'realistic' mathematics problems: Just how realistic are children ready to be? *Educational Studies in Mathematics, 49*, 1-23.

Coquin-Veinnot, D., & Moreau, S. (2007). Arithmetic problems at school: When there is an apparent contradiction between the situation model and the problem model. *British Journal of Educational Psychology, 77*, 69-80. doi:10.1348/000709905X79121

Cvencek, D., Meltzoff, A. N., & Greenwalk, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development, 0*(0), 1-14. doi:10.1111/j.1467-8624.2010.01.529x

Dark, V. J., & Benbow, C. P. (1990). Enhanced problem translation and short-term memory: Components of mathematical talent. *Journal of Educational Psychology, 82*, 420-429.

De Corte, E., Verschaffel, L., & Lasure, S. (1995, April). *Word problems: Game or reality? Studies of children's beliefs about the role of real-world knowledge in mathematical modeling.* Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, US.

DeFranco, T. C., Curcio, F. R. (1993). A division problems with a remainder embedded across two contexts: Children's solutions in restrictive vs. real-world settings. *Focus on Learning Problems in Mathematics, 19*, 58-72.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence, 32*, 25-32. doi:10.1016/S0160-2896(03)00061-8

Dewolf, T., Van Dooren, W., Ev Cimen, E., & Verschaffel, L. (2014). The impact of illustrations and warnings on solving mathematical word problems realistically. *The Journal of Experimental Education 82*, 103-120. doi:10.1080/00220973.2012.745468

Dewolf, T., Van Dooren, W., Hermens, F., & Verschaffel, L. (2015). Do students attend to representational illustrations of non-standard mathematical word problems, and, if so, how helpful are they? *Instructional Science, 43*, 147-171. doi:10.1007/s11251-014-9332-7

Dewolf, T, Van Dooren, W., & Verschaffel, L. (2016). Can visual aids in representational illustrations help pupils to solve mathematical word problems more realistically? *European Journal of Psychology Education,* 1-17. doi:10.1007/s10212-016-0308-7

Dewolf, T., Van Dooren, W., & Verschaffel, L. (2011). Upper elementary school children's understanding and solution of a quantitative problem inside and outside the mathematics class. *Learning and Instruction, 21*, 770-780. doi:10.1016/j.learninstruc.2011.05.003

Friedman, L. (1995). The space factor in mathematics: Gender differences. *Review of Educational Research, 65*, 22-50. doi:10.3102/00346543065001022

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., …Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29-43. doi:10.1037/002-0663.98.1.29

Gravemeijer, K. (1997). Commentary solving word problems: A case of modeling? *Learning and Instruction, 7*, 389-397.

Greer, B. (1993). The mathematical modeling perspective on wor(l)d problems. *Journal of Mathematical Behavior, 12*, 239-250.

Greer, B., Verschaffel, L., & De Corte, E. (2003). "The answer is really 4.5": Beliefs about word problems. In G. Leder, E. Pehkonen, & G. Torner (Eds.), *Beliefs: A hidden variable in mathematics education* (pp. 271-292). Secaucus, NJ, US: Kluwer Academic Publishers. Retrieved from http://www.ebrary.com

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139-155. doi:10.1037/0033-2909.107.2.139

Inoue, N. (2005). The realistic reasons behind unrealistic solutions: The role of interpretive activity in word problem solving. *Learning and Instruction, 15*, 69-83. doi:10.1016/j.learninstruc.2004.12.004

Inoue, N. (2008). Minimalism as a guiding principle: Linking mathematical learning to everyday knowledge. *Mathematical Thinking and Learning, 10*, 36-67. doi:10.1080/10986060701820293

Jonassen, D. H. (2003). Designing research-based instruction for story problems. *Educational Psychology Review, 15*, 267-296. doi:1040-726X/03/0900-0267/0

Johnson, E. S. (1984). Sex differences in problem solving. *Journal of Educational Psychology, 76*(6), 1359-1371.

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

Kaizer, C., & Shore, B. M. (1995). Strategy flexibility in more and less competent students on mathematical word problems. *Creativity Research Journal, 8*(1), 77-82. https://doi.org/10.1207/s15326934crj0801_6

Kyttälä, M., & Björn, P. M. (2014). The role of literacy skills in adolescents' mathematics word problems performance: Controlling for visuo-spatial ability and mathematics anxiety. *Learning and Individual Differences, 29*, 59-66. doi:10.1016/j.lindif.2013.10.010

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. doi:10.1016/j.jecp.2004.07.001

Lummis, M., & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology, 26*(2), 254-261.

Martin, S. A., & Bassok, M. (2005). Effects of semantic cues on mathematical modeling: Evidence from word-problem solving and equation construction tasks. *Memory & Cognition, 33*(3), 471-478.

Matalliotaki, E. (2012). Resolution of division problems by young children: What are children capable of and under what conditions? *European Early Childhood Education Research Journal, 20*(2), 283-299. doi:10.1080/1350293X.2012.681132

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

Muth, K. D. (1992). Brief research report: Extraneous information and extra steps in arithmetic word problems. *Contemporary Educational Psychology, 17*, 278-285.

Muth, K. D., & Glynn, S. M. (1985). Comprehension of arithmetic word problems by fifth-grade pupils: Representations and selection of information. *British Journal of Educational Psychology, 73*, 109-121.

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 7*, 37-58. doi:10.1007/s10649-007-9083-3

Palm, T., & Nyström, P. (2009). Gender aspects of sense making in word problem solving. *Journal of Mathematical Modelling and Application, 1*, 59-76.

Pape, S. J. (2004). Middle school children's problem-solving behavior: A cognitive analysis from a reading comprehension perspective. *Journal for Research in Mathematics Education, 35*, 187-219.

Pimperton, H., & Nation, K. (2010). Understanding words, understanding numbers: An exploration of the mathematical profiles of poor comprehenders. *British Journal of Educational Psychology, 80*, 255-268. doi:10.1348/000709909X477251

Reusser, K., & Stebler, R. (1997). Every word problem has a solution – The social rationality of mathematical modeling in schools. *Learning and Instructions, 7*, 309-327.

Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce on formal and informal mathematics. In *Informal Reasoning and Education.* J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.) (pp. 311-343). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Silver, E. A., Shapiro, L. J., & Deutsch, A. (1993). Sense making and the solution of division problems involving remainders: An Examination of middle school students' solution processes and their interpretations of solutions. *Journal for Research in Mathematics Education, 24*, 117-135.

Swanson, H. L. (2004). Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Memory & Cognition, 32*, 648-661.

Thevenot, C., Devidal, M., Barrouillet, P., & Fayol, M. (2007). Why does placing the question before an arithmetic word problem improve performance? A situation model account. *The Quarterly Journal of Experimental Psychology, 60*(1), 43-56. doi:10.1080/17470210600587927

U.S. Department of Education. (2008). *Foundations for success: The final report of the national mathematics advisory panel* (ED04CO0082/0001).

Verschaffel, L., & De Corte, E. (1997a). Word problems: A vehicle for promoting Kailauthentic mathematical understanding and problems solving in the primary school? In *Learning and Teaching Mathematics.* T. Nunes, & P. Bryant (Eds.) (pp. 69-97). East Sussex, UK: Psychology Press Ltd.

Verschaffel, L., & De Corte, E. (1997b). Teaching realistic mathematical modeling in the elementary school: A teaching experiment with fifth graders. *Journal for Research in Mathematics Education, 28*, 577-601.

Verschaffel, L., De Corte, E., & Borghart, I. (1997). Preservice teachers' conceptions and beliefs about the role of real-world knowledge in mathematical modeling of school word problems. *Learning and Instruction, 7*, 339-359.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4*, 273-294.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making Sense of Word Problems.* Lisse, The Netherlands: Swets & Zeitlinger.

Verschaffel, L., Van Dooren, W., Greer, B., & Mukhopadhyay, S. (2010). Reconceptualising word problems as exercises in mathematical modeling. *Journal für Mathematik-Didaktik, 31*, 3-29. doi:10.1007/s13138-010-0007-x

Vilenius-Tuohimaa, P., Aunola, K., & Nurmi, J-E. (2008). The association between mathematical word problems and reading comprehension. *Educational Journal, 28,* 409-426. doi:10.1080/01443410701708228

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

Weyns, A., Van Dooren, W., Dewolf, T., & Verschaffel, L. (2016). The effect of emphasizing the realistic modelling complexity in the text or picture on pupils' realistic solution of P-items. *An International Journal of Experimental Educational Psychology,* doi:10.1080/01443410.2016.1259461

Wyndhamn, J., & Säljö, R. (1997). Word problems and mathematical reasoning – A study of children's mastery of reference and meaning in textual realities. *Learning and Instruction, 7*, 361-382.

Xin, A., Lin, C., Zhang, L., Yan, R. (2007). The performance of Chinese primary school students on realistic arithmetic word problems. *Educational Psychology in*

Practice: Theory, Research, and Practice in Educational Psychology, 23, 145-159. doi:10.1080/02667360701320853

Xin, A., & Zhang, L. (2009). Cognitive holding power, fluid intelligence, and mathematical achievement as predictors of children's realistic problem solving. *Learning and Individual Differences, 19*, 124-129. doi:10.1016/j.lindif.2008.05.006

Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology, 110*, 481-498. doi:10.1016/j.jecp.2011.06.001

Zwann, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.

Chapter 2

Abstract

Researchers investigating how children can incorporate real world knowledge into mathematical word problems have largely focused on ways to improve realistic thinking. Considerably less research has examined how realistic responding in these problems is associated with individual differences. This study tested whether general academic abilities are related to realistic responses, and whether realistic responding is related to general word problem solving. In our sample of sixth-grade students, general math and verbal academic abilities were not independently predictive of solving realistic word problems; however, performance on realistic word problems was independently predictive of solving basic word problems. As such, realistic word problems may reflect problem solving ability independent of general academic ability, and therefore is an ability worth fostering.

Keywords: word problems, academic abilities, educational psychology, math cognition

The past 30 years has seen a growing interest in students' use of real world knowledge when solving mathematical word problems (for a detailed summary see Verschaffel, Greer, & De Corte, 2000). In these problems, students have to consider that the sprinting speed of someone running 100 m cannot be maintained over a kilometer, or that a child cannot receive a half of a balloon in a sharing task, or that tying separate pieces of rope together requires extra length to account for the knots (De Corte, Verschaffel, & Lasure, 1995; Verschaffel, De Corte, & Lasure, 1994; Yoshida, Verschaffel, & De Corte, 1997). The findings from this body of research provide a general consensus that elementary and middle school students, undergraduate students, and preservice teachers have a strong tendency to disregard real world knowledge when solving mathematical word problems (WPs) that require realistic consideration (Inoue, 2005, 2008; Verschaffel, De Corte, & Borghart, 1997; Verschaffel et al., 2000). This trend is similar in European (De Corte et al., 1995) and Asian (Inoue, 2005, 2008; Yoshida et al., 1997) countries.

Most of the research on this phenomenon involves attempts to improve students' performance. The experimental methodologies used to improve realistic responding have included more detail in the problem text (DeFranco & Curcio, 1993; Palm, 2008; Palm & Nyström, 2009), including informative pictures (Dewolf, Van Dooren, Ev Cimen, & Verschaffel, 2014; Dewolf, Van Dooren, & Verschaffel, 2016; Weyns, Van Dooren, Dewolf, & Verschaffel, 2016), and warnings (Reusser & Stebler, 1997; Xin, Lin, Zhang, & Yan, 2007; Yoshida et al., 1997). The warnings have been provided orally (Dewolf et al., 2014; Xin et al., 2007) and written (Dewolf et al., 2014; Reusser & Stebler, 1997; Yoshida et al., 1997), both at the beginning of test booklets (Dewolf et al., 2014; Yoshida et al., 1997), just before the WP (Dewolf et al., 2014; Reusser & Stebler, 1997), and embedded within the problem itself (Reusser & Stebler, 1997; Weyns et al., 2016). Despite the best efforts of researchers, the methodological modifications have done little to increase the number of realistic responses (see DeFranco & Curcio, 1993; Palm, 2008; Palm & Nyström, 2009). When improvements have been found, students provide realistic responses one fifth to one half of the time (Palm, 2008; Palm & Nyström, 2009).

A key question overlooked in this line of literature – and arguably a more important question than how to improve realistic responding – is: Why should we trying to improve realistic responding? Although it may seem intuitive that students' ability to take real world information into account would reflect well on student's general problem-solving ability, and demonstrate good conceptual understanding, this may not necessarily be true. Students' experience of word problems in schools is often called the 'suspension of sense making' (Schoenfeld, 1991; Verschaffel & De Corte, 1997a), where the rules of the 'game of word problems' (Verschaffel & De Corte, 1985, 1997a) say you should ignore real world considerations. Given all this experience in the classroom, it is difficult to fault students for ignoring real-world information in word problems. Those who do pay attention to it may either be slow to pick up on the fact that these word problems are just excuses to practice the math topic just learned, or students are too easily distracted by "irrelevant" information. As such, it is worth asking whether the ability to apply realistic reasoning to situations in math WPs is a useful skill, and whether learning how to apply this type of reasoning to word problems benefits general math ability. The purpose of the study reported here is to test whether realistic responding is related to other general academic abilities (e.g. verbal ability, math reasoning) that are related to success on standard WPs, and also to test whether realistic responding itself is related to performance on standard WPs. If realistic responding represents a valuable insight into problem solving, and not a failure to understand that suspension of belief is expected, then general academic abilities should be related to realistic word problem success. Furthermore, realistic responding should be related to success on standard word problems, when "success" is judged by answers that take reality into account.

Based on the findings in the literature, we know it is extremely challenging to improve students' realistic responding. However, very little research has examined individual differences in realistic reasoning. There is much more work on this topic when researchers have evaluated individual differences with standard WPs than with realistic WPs. As might be expected, skills such as prior mathematical ability (Kail & Hall, 1999; Muth, 1984; Xin & Zhang, 2009) and verbal ability (Bjork & Bowyer-Crane, 2013; Muth, 1984; Pimperton & Nation, 2010; Vilenius-Tuohimaa, Aunola, & Nurmi, 2008) have consistently been found to predict standard word problem performance.

Muth (1984) and Fuchs et al. (2006) have shown, in sixth- and third-graders, a positive correlation between arithmetic performance and math WP performance [$r_{(198)}$ = .60 and $r_{(310)}$ = 0.45, respectively]. Muth (1984) found that, once reading ability was controlled for, computation ability continued to predict 7.62% of the variance in WPs performance. Fuchs et al. (2006) used path analysis and found that once phonological decoding, sight-word efficiency, and language were all accounted for; arithmetic skill remained an independent predictor of math WP performance. The findings of Muth (1984) and Fuchs et al. (2006) indicate the importance of arithmetic as being a foundational skill for solving arithmetic WPs.

Despite this strong relation between arithmetic ability and standard word problem performance, research has also demonstrated that students feel more comfortable with and perform better when solving arithmetic calculation problems in comparison to arithmetic WPs (Muth, 1984). For example, Bjork and Bowyer-Crane (2013) had 6- and 7-year-olds complete 15 arithmetic calculations and 15 WPs that had the exact same numbers and operations as were used in the arithmetic calculations (i.e., the calculation problem '9 + 5 =' corresponded to the matched WP, 'Ann has 9 pennies and her friend has 5 pennies. If they add their pennies together, how many do they have?' p. 1351). The students performed significantly better on the arithmetic calculation problems than the arithmetic WPs.

Perhaps it is the verbiage of word problems that makes them more difficult. As it turns out, verbal ability predicts success in solving standard WPs. Bjork and Bowyer-Crane (2013), also had the 6- and 7-year-old children complete a series of tasks examining different cognitive abilities using the Wechsler Abbreviated Scale of Intelligence (e.g., phonological awareness, verbal ability, reading accuracy, and reading comprehension). They wanted to determine which cognitive skills were important predictors for success on numerical operation and word problems. They found that only phonological awareness was a predictor of numerical operation performance, while phonological awareness, reading comprehension, and numerical operations were all unique predictors of performance in the WP task. Bjork and Bowyer-Crane argue that these findings support the importance of reading comprehension as a cognitive skill necessary for solving math WPs.

With a sample of older students (9- to 10-year-olds), Vilenius-Tuohimaa et al. (2008) categorized students as good or poor readers based on their technical reading skills as measured by a Finnish reading test known as the ALLU (Ala-Asteen Lukutesti) Reading Test for Primary School. They found that children in the good reading group performed significantly better on a math WPs task than the students in the poor reading group. Using path analysis, the authors also found that technical reading scores accounted for 18% of the variance in scores on the math WP task, and that, after technical reading skills were controlled for, reading comprehension skills continued to predict variance in math WPs. Vilenius-Tuohimaa et al. suggest both reading comprehension and math WP performance involve a reasoning component. Even in the older grades (e.g., eighth-grade) literacy skills continue to predict math WP performance (Kyttälä & Björn, 2014). However, these findings have not been applied to realistic responses in WPs that require the use of real-world knowledge. One would assume the same skills should be relevant to solving realistic WPs, but they have yet to be explicitly tested.

While there appears to be a substantial amount of literature on the factors that contribute to success in standard mathematical WPs, there is considerably less information on realistic WPs. One exception is Xin and Zhang (2009), who gave fourth- and sixth-grade students modified realistic WPs from Verschaffel et al. (1994). They found that prior mathematical achievement (as measured by students' previous semester exam scores) and fluid intelligence (as measured by Raven's Matrices – modified for Chinese students) to be significant predictors of success on both standard and realistic WPs. However, Xin and Zhang tested only for previous math performance, and did not test for specific academic abilities (e.g., math calculation vs. math reasoning). No other researchers, including Xin and Zhang (2009), have tested whether verbal ability predicts realistic responding, or if the relation between realistic word problems and standard word problems still persists after controlling for general academic abilities. Furthermore, Xin and Zhang (2009) used a Chinese sample, and it would worthwhile to see if their general finding would also be true in a North American sample.

The literature investigating realistic problem solving has largely focused on trying to improve realistic responding and, in doing so, has implicitly assumed that this is an important skill. If that is true, then it seems reasonable that the factors that predict successfully solving a basic math WP would also be predictors of successfully applying real-world knowledge to WPs that require the use of real-world consideration. The current study tested this assumption in two ways. First, a range of academic abilities (e.g., different aspects of mathematical and verbal ability) were tested to determine if they are related to realistic word-problem solving in same way they are related to standard word-problem solving. Second, realistic word problem ability and standard word problem ability were tested to determine if the two abilities were still related to each other after controlling for the general academic abilities.

Method

Participants

Eighty-five grade six students (41 boys and 44 girls; M_{Age} = 11.737 yrs, SD = .359) participated from six different elementary schools in a mid-size Canadian city. Information on ethnicity was not collected; however, the city largely consists of a Caucasian population. The participating schools included those from low socioeconomic areas to those in the highest socioeconomic areas. As these children are below the age of consent, parental consent forms were collected prior to data collection (Appendix A). Two participants were removed from the data analyses because of extreme scores ($z < -3.00$) on the Raven's Matrices and the WJ III Ach composite scores, as well as 0% accuracy on the standard WPs. One other student was an outlier on the Raven's Matrices ($z < -3.00$) and was removed from analyses using a case-wise selection method, and 10 students did not complete all measures of the WJ III Ach. A total of 82 students were used in most analyses (40 boys and 42 girls; M_{Age} = 11.73 yrs, *SD* = .362), and 72 were used in the analyses involving the WJ III Ach (32 boys and 40 girls; M_{Age} = 11.71 yrs, *SD* = .363).

Materials

Demographic Form. Demographic information was acquired, such as, the child's birthday, sex, their interests and hobbies, and any extracurricular activities in which they participate (see Appendix B). This information was used to assist with ensuring the WPs were not biased towards any particular student's interests as context has been found to be important in gender differences (Boaler, 1993a, 1994; Palm & Nyström, 2009)].

Raven's Standard Progressive Matrices (Raven, Raven, & Court, 1998). The Raven's Matrices is a standardized non-verbal, multiple-choice reasoning task used to assess problem solving ability, general intelligence, and cognitive ability (fluid intelligence). Raven and colleagues report a split-half internal consistency modal value of .91. The Raven's was designed to be useful for individuals of all ages and is made up of 60 problems – five series of 12 problems each – of diagrammatic puzzles with a part missing, which the test taker must find among the options provided. The initial problems are easy but the problems increase in difficulty as the child progresses through the test.

Participants received a smaller selection of 28 items chosen based on age norms deemed appropriate for grade 6 children. Age norms range from children five and a half years-of-age to adults 85 years-of-age, which make it appropriate for the children to be sampled in this study. The Raven's is a group administered assessment.

Woodcock-Johnson III Tests of Achievement (WJ III Ach; Woodcock, McGrew, & Mather, 2001). The WJ III Ach is a multidimensional test of general intelligence for ages 2-90 years. The measure includes 22 tests for measuring skills in reading, mathematics, and writing, as well as important oral language abilities and academic knowledge, each of which measure one or more cognitive processes. Subsections of this measure were used for the current study, specifically those measuring reading comprehension and mathematical knowledge and ability. This standardized measure has undergone rigorous testing to ensure reliability and validity. All substests were administered one-on-one.

Calculation. Calculation is a test of math achievement measuring the ability to perform mathematical computations such as addition, subtraction, multiplication, division, and combinations of these basic operations, as well as some geometric, trigonometric, logarithmic, and calculus operations. The starting point was determined based on an estimate of the participant's present level of computation skill. The estimate is determined by the test makers and the starting point for grade 6 students is item 9. Calculation has a median reliability of .85 in the range of 5-19 year olds.

Math Fluency. Math Fluency measured accuracy on simple addition, subtraction, and multiplication facts. Participants were presented with a series of problems and were asked to complete as many as possible in the 3-min time limit. Math Fluency has a median reliability of .89 in the range of 5-19 year olds.

Passage Comprehension. Passage Comprehension began by examining symbolic learning. For example, the student would see an image of a house with a stick figure beside it and a statement with a word missing: “The house is _____ than the man”. The student had to complete the statement so it described the image. Participants eventually moved onto reading short passages and identifying missing key words that made sense in the context of the passage. The items increased in difficulty by the removal of the pictorial stimuli and by increasing passage length, level of vocabulary, and complexity of syntactic and semantic cues. Passage Comprehension has a median reliability of .83 in the range of 5-19 year olds.

Applied Problems. Applied Problems required participants to analyze (e.g., determine the necessary math operation) and solve math problems. Participants must listen to the problem, recognize the procedure to be followed and then perform simple calculations. Each problem was visible to students but the experimenter read the problem aloud. As items can include extraneous information the person must identify not only the operation necessary to solve the problem but also the relevant information. Items increased in difficulty with more complex calculations. Applied Problems has a median reliability of .92 in the range of 5-19 year olds.

Reading Vocabulary. Reading Vocabulary measured skill in reading words and providing appropriate meanings. This test included three subtests: Synonyms, Antonyms, and Analogies. Synonyms required reading a word and providing a word that means the same, Antonyms required reading a word and providing a word that means the opposite, and Analogies required reading three words of an analogy and providing the fourth word to complete the analogy. Items increased in difficulty for each subtest. Reading Vocabulary has a median reliability of .87 in the range of 5-19 year olds.

Quantitative Concepts. Quantitative Concepts measured knowledge of mathematical concepts, symbols, and vocabulary. This test included two subtests: Concepts and Number Series. Concepts began with requiring the person to count and identify numbers, shapes, sequences, and further items required knowledge of mathematical terms and formulas. Number Series required the person to look at a series of numbers, figure out the pattern, and then provide the missing number in the series. Quantitative Concepts has a median reliability of .90 in the range of 5-19 year olds.

Realistic and Standard Mathematical Word Problems (Greer, 1993; Verschaffel et al., 1994). Realistic and standard WPs were taken from Greer (1993) and Verschaffel et al. (1994), as these problems have been used repeatedly in the literature. Each student received a pencil-and-paper task with 10 problems, 5 standard (S-) and 5 realistic (R-) items, the realistic items are those that require consideration of realistic knowledge to answer (see Appendix C). Four versions of this task were used to ensure all standard-items (10) and realistic-items (10) were answered and that the WPs were not always seen in the same order. This task included an area for an answer and a section for comments (see Appendix D). Students were not provided with assistance during this task and were asked to complete all problems on their own. This assessment was group administered and students were given an hour to complete all 10 word problems.

Final Interview (Palm, 2008). The aim of the interview was to investigate whether students had given any realistic considerations to the situation outlined in the

WP but failed to use those considerations when providing a solution, or failed to report them in the comments section. The interviews were conducted one-on-one and students were shown their original non-realistic reaction answers to realistic-items. The first two questions were designed to detect if real-world considerations were made during the process of solving the WP. If no real-world considerations were reported based on the first two questions, the student was asked the third question, which pointed out a specific realistic feature of the WP. If the child did consider the realistic nature of the WP they were asked a fourth question, specifically, why they did not use that information to answer the problem. Finally, if the student did not recognize the difficulty in the WP they were asked the last question, which explicitly pointed out the realistic nature of the realistic-item. For example, in the problem "John's best time to run 100m is 17 sec. How long will it take to run 1 km?" (Verschaffel et al., 1994, p. 276) the student was asked if they thought the runner could keep the same speed during the whole race (see Appendix E for interview questions).

Audio Recording Device. An audio recording device was used to record the interview sessions. This assisted the researcher in later transcription and coding of realistic considerations provided by students. Recordings from the interview sessions were used so that inter-rater reliability could be established.

Procedure

The experimental and control conditions were randomly assigned to school. The results of this manipulation are not discussed here, as they were not related to the hypotheses of the current study. Data collection occurred across multiple sessions, beginning with the Raven's Matrices (Raven et al., 1998) and WP booklets. Students were given the same, spoken aloud, group instructions for the Raven's Matrices, and two practice items were completed with the researcher (Raven et al., 1998). After this task, students were asked to complete 10 WPs; instructions were those used by Verschaffel et al., (1994).

Following the written assessments, the WJ III Ach measures were administered in individual interviews. Students were randomly assigned to be tested on Calculation, Passage Comprehension, and Applied Problems first or on Math Fluency, Reading Vocabulary, and Quantitative Concepts first. As the WJ III Ach measure is standardized, test administration followed the existing protocol. Each testing session lasted 30 to 40 min. Individual subtest and composite cores on the WJ III Ach were calculated using the computer software Compuscore and Profiles Program (Version 3.0, Schrank, & Woodcock, 2007). Three composite scores are reported: Math Calculation Skills includes Math Calculation and Math Fluency, and Math Reasoning combines Applied Problems and Quantitative Concepts, and Reading Comprehension combines Reading Vocabulary and Passage Comprehension.

Results

The goal of the preliminary analysis was to determine whether students' ability to use their real-world knowledge for solving problematic mathematical word problems is comparable to the results found in previous research (e.g., 16-17%, Verschaffel et al., 2000). In the current data, the realistic response (RR) rate was 21.5%, 95% CI [15.3%, 27.7%], and the response rate in the literature falls within this confidence interval. As such, our sample is comparable to participants in previous research.

The first main analysis compared measures of individual differences to realistic responses on realistic WPs and accuracy on standard WPs. These correlations can be seen in Table 1. All measures of individual differences were positively correlated with one another and with accuracy on the standard WPs. The pattern is not quite as strong with realistic responses, as the Raven's was not correlated with it and Math Calculation was only marginally related. The strongest correlation was with Reading Comprehension.

Table 1

Pearson's Correlations for Realistic Responses, Raven's Matrices, WJ III Ach Composite Measures, and Standard Word Problems

	Realistic Responses	Raven's Matrices	Reading Comprehension	Math Reasoning	Math Calculation
Realistic Responses	-				
Raven's Matrices	.118 [b]	-			
Reading Comprehension	.308** [a]	.288* [a]	-		
Math Reasoning	.279* [a]	.418** [a]	.541** [a]	-	
Math Calculation	.230† [a]	.246** [a]	.356** [a]	.744** [a]	-
Standard WPs	.358** [b]	.394** [b]	.464** [a]	.672** [a]	.528** [a]

Note. $N^a = 72$, $N^b = 82$, †$p < .10$, *$p < .05$, **$p < .01$

The next analysis was to determine whether these academic abilities would predict unique variance after controlling for the other abilities. Composite measures of the WJ III Ach, (Reading Comprehension and Math Calculation Skills) were expected to be positive predictors of realistic responses on the realistic-items based on the literature on standard WPs. In the regression analysis, the total number of RRs was entered as the criterion variable and the three composite scores of the WJ III Ach, Gender, and Raven's Matrices scores were entered as predictor variables. The three composite measures, Gender, and Raven's Matrices scores did not account for a significant amount of variance in RRs, $F(5, 66) = 1.961$, $p = .096$ (Table 2). The results do not support our hypothesis that Gender and Raven's Matrices scores would account for a significant amount of variance in RRs, and the Reading Comprehension composite measure only approached significance.

Table 2

Predictors of Total Number of Realistic Responses

Variable	R^2	*B*	*SE (B)*	β
Constant		-1.577	1.285	
Raven's Matrices		.002	.035	.007
Reading Comprehension		.024†	.013	.251
Math Calculation		.001	.013	.016
Math Reasoning		.008	.017	.093
Gender		-.246	.237	-.134
Total Model	.129			

Note. *B* denotes the variable estimate, *SE B* denotes the standard errors of the variable estimate, β denotes the standardized estimate, and †$p < .10$

The last analyses tested the relation between realistic responding and standard WP performance while controlling for general academic abilities. Using hierarchical regression analysis, performance on the standard WP items (S-items) was entered as the criterion. Scores on the Raven's Matrices, all three composite measures of the WJ III Ach, and Gender were entered into the first block, and RRs was entered in the second block. Table 3 shows that when all predictors were entered into the model together they account for a statistically significant amount of the variance in accuracy on S-items, $F(6, 65) = 12.783$, $p < .0005$. The data also show that, after all individual differences measures were entered into the model, RRs still accounted for variability in performance on the S-items, $\Delta R^2 = .050$, $F(1, 65) = 7.132$, $p = .010$. Math Reasoning was also found to be a unique predictor of variance in performance on S-items, accounting for 6.2% of the variance. In other words, realistic responding predicted unique variance in students' performance on standard WPs even after controlling for general academic abilities.

Table 3

Predictors of Accuracy on Standard WPs

Variable	R^2	*B*	*SE (B)*	β
Constant		-5.383	1.276	
Raven's Matrices		.057	.034	.157
Reading Comprehension		.009	.013	.068
Math Calculation		.011	.013	.112
Math Reasoning		.048**	.016	.437
Gender		.165	.234	.067
RRs		.323*	.121	.240
Total Model	.541**			

Note. *B* denotes the variable estimate, *SE B* denotes the standard errors of the variable estimate, β denotes the standardized estimate, and $*p < .05$, $**p < .01$

Discussion

The primary goal of the current work was to examine individual differences in using realistic knowledge to solve realistic WPs. A secondary goal of the current work was to determine whether performance on realistic word problems is predictive of performance on standard word problems. Measures of fluid intelligence (e.g., Raven's Standard Progressive Matrices), achievement (e.g., Reading Comprehension, Math Calculation Skills, and Math Reasoning), and gender were used as indices of individual differences.

The first question to be examined was whether the academic abilities that have been found to be related to success on basic WPs are also related to success on realistic WPs. In the existing literature, both verbal ability (Bjork & Bowyer-Crane, 2013;

Pimperton & Nation, 2010; Vilenius-Tuohimma et al., 2008) and prior math ability (Kail & Hall, 1999; Muth, 1984) have been found to be predictors of performance on standard WPs (Fuchs et al., 2006; Muth, 1984). Except for one study, most of this research was done using standard WPs (Xin & Zhang, 2009). There was no reason to assume that the same academic abilities would not be related to performance on realistic WPs, and this study aimed to extend the findings of Xin and Zhang (2009) to North American students.

The findings from the current study do not support the notion that the academic abilities related to success on standard WPs are the same as those related to success on realistic WPs. While there was no evidence of individual academic abilities being correlated with performance on realistic WPs, once all factors were entered into a regression model together, they did not account for a significant amount of variability and no one predictor accounted for unique variance. This study was also unable to extend the findings of Xin and Zhang (2009) with North American students. Our data failed to indicate that math calculation or fluid intelligence predicts the frequency of realistic responses given to realistic WPs. Likewise, the data also failed to show that math calculation ability predicts success on standard WPs. This is especially surprising given the evidence in the existing literature which does support this point (Kail & Hall, 1999; Muth, 1984; Xin & Zhang, 2009). An unexpected finding in the current study was that math reasoning was a skill that positively predicted performance on standard WPs while reading comprehension and math calculation ability did not. Not only was the data unable to show that academic abilities which predict standard WP solving performance are also able to predict realistic WP solving performance, the data did not support these academic abilities as being related to standard WP solving performance. It may be that performance on realistic WPs is better accounted for by variables other than academic abilities. Perhaps there are characteristics such as a desire for understanding complex problems or enjoying thinking about abstract ideas which are better predictors of performance on realistic WPs. These kinds of attributes are not readily measured by the same assessments that are designed to measure academic abilities. It is possible that there are personality traits that may predict performance on realistic WPs. This question should be examined in future research.

Initially, it was surprising that reading comprehension was not a unique predictor of realistic responding, especially given the extant evidence indicating reading comprehension is an important factor for success on standard WPs (Bjork & Bowyer-Crane, 2013; Muth, 1984; Pimperton & Nation, 2010; Vilenius-Tuohimma et al., 2008). However, the research on standard WPs has demonstrated that poor reading comprehension may be associated with poor math reasoning (Pimperton & Nation, 2010; Vilenius-Tuohimaa et al., 2008). It is possible that reading comprehension was not a significant predictor of realistic responding because it is tied up in math reasoning skills and therefore does not account for unique variance on its own. Considering the highly-significant correlation (see Table 1) between reading comprehension and math reasoning, this is a plausible explanation.

The major finding in the current study, and debatably the more important finding, was how RRs relates to standard WP performance. That RRs continue to predict variance in the standard WPs after controlling for measures of fluid intelligence, achievement, and gender suggests there is something about being more likely to incorporate real world knowledge into the solution process that makes accuracy in standard WPs more likely, even after a variety of individual factors have been accounted for. This begs the question: What is it about recognizing and implementing real-world knowledge into the solution process of a mathematical WP that helps with general WP performance? One explanation could be the usefulness of creating a situational model (see Zwann & Radvansky, 1998). It could be that the students who are able to construct a situation model for a problem that involves realistic consideration are also benefitting from building situation models for standard WPs. In other words, if students can consider practical constraints and complexities in their reasoning during realistic word problem solving, then they are well able to construct a situation model to conceptualize the less demanding standard word problem paradigms. This finding also suggests that knowing how to use realistic knowledge may not get in the way of knowing how to “play the game” of word problems. These students have developed a skill in understanding the realistic components, and they may know when and when not to use it.

Our initial hypotheses were that: a) incorporating real-world knowledge into problems that require it should be related to general academic abilities in the same way that standard WP performance is; and b) that realistic word problems should predict unique variance in standard WP performance. The fact that our first hypothesis was not supported but the second one was, has interesting implications. It was proposed that realistic responding reflects a general problem-solving ability that is independent of general academic ability. Our standard WPs are comparable to mathematical textbook problems (e.g., situation makes sense, straight forward application of basic arithmetic, no extraneous information, and one possible answer; see Reusser, 1988; Verschaffel et al., 2000) which means that having realistic knowledge makes one better at solving the types of word problems that are presented throughout one's primary and secondary mathematical education. These results offer proper support for the general assumption in the literature that realistic responding is a skill worth investigating and fostering.

References

Bjork, I. A., & Bowyer-Crane, C. (2013). Cognitive skills used to solve mathematical word problems and numerical operations: A study of 6- to 7-year-old children. *European Journal of Psychological Education, 28*, 1345-1360. doi:10.1007/s10212-012-0169-7

De Corte, E., Verschaffel, L., & Lasure, S. (1995, April). *Word problems: Game or reality? Studies of children's beliefs about the role of real-world knowledge in mathematical modeling.* Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, US.

DeFranco, T. C., Curcio, F. R. (1993). A division problems with a remainder embedded across two contexts: Children's solutions in restrictive vs. real-world settings. *Focus on Learning Problems in Mathematics, 19*, 58-72.

Dewolf, T., Van Dooren, W., Ev Cimen, E., & Verschaffel, L. (2014). The impact of illustrations and warnings on solving mathematical word problems realistically. *The Journal of Experimental Education, 82*, 103-120. doi:10.1080/00220973.2012.745468

Dewolf, T., Van Dooren, W., & Verschaffel, L. (2016). Can visual aids in representational illustrations help pupils to solve mathematical word problems more realistically? *European Journal of Psychology Education,* 1-17. doi:10.1007/s10212-016-0308-7

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., …Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29-43. doi:10.1037/002-0663.98.1.29

Greer, B. (1993). The mathematical modeling perspective on wor(l)d problems. *Journal of Mathematical Behavior, 12*, 239-250.

Inoue, N. (2005). The realistic reasons behind unrealistic solutions: The role of interpretive activity in word problem solving. *Learning and Instruction, 15*, 69-83. doi:10.1016/j.learninstruc.2004.12.004

Inoue, N. (2008). Minimalism as a guiding principle: Linking mathematical learning to everyday knowledge. *Mathematical Thinking and Learning, 10*, 36-67. doi:10.1080/10986060701820293

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

Kyttälä, M., & Björn, P. M. (2014). The role of literacy skills in adolescents' mathematics word problems performance: Controlling for visuo-spatial ability and mathematics anxiety. *Learning and Individual Differences, 29*, 59-66. doi:10.1016/j.lindif.2013.10.010

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

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 7*, 37-58. doi:10.1007/s10649-007-9083-3

Palm, T., & Nyström, P. (2009). Gender aspects of sense making in word problem solving. *Journal of Mathematical Modelling and Application, 1*, 59-76.

Pimperton, H., & Nation, K. (2010). Understanding words, understanding numbers: An exploration of the mathematical profiles of poor comprehenders. *British Journal of Educational Psychology, 80*, 255-268. doi:10.1348/000709909X477251

Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard Progressive Matrices.* Texas, US: Harcourt.

Reusser, K. (1988). Problem solving beyond the logic of things: Contextual effect on understanding and solving word problems. *Instructional Science, 17*, 309-338.

Reusser, K., & Stebler, R. (1997). Every word problem has a solution – The social rationality of mathematical modeling in schools. *Learning and Instruction, 7*, 309-327.

Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce on formal and informal mathematics. In *Informal Reasoning and Education.* J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.) (pp. 311-343). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Schrank, F. A., & Woodcock, R. W. (2001). WJ III Compuscore and Profiles Program [Computer Software]. *Woodcock-Johnson III.* Rolling Meadows, IL: Riverside.

Verschaffel, L., & De Corte, E. (1997a). Word problems: A vehicle for promoting authentic mathematical understanding and problem solving in the primary school? In *Learning and Teaching Mathematics.* T. Nuñes, & P. Bryant (Eds.) (pp. 69-97). East Sussex, UK: Psychology Press Ltd.

Verschaffel, L., De Corte, E., & Borghart, I. (1997). Preservice teachers' conceptions and beliefs about the role of real-world knowledge in mathematical modeling of school word problems. *Learning and Instruction, 7*, 339-359.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4*, 27.-294.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making Sense of Word Problems.* Lisse, The Netherlands: Swets & Zeitlinger.

Vilenius-Tuohimaa, P., Aunola, K., & Nurmi, J-E. (2008). The association between mathematical word problems and reading comprehension. *Educational Journal, 28,* 409-426. doi:10.1080/01443410701708228

Weyns, A., Van Dooren, W., Dewolf, T., & Verschaffel, L. (2016). The effect of emphasizing the realistic modelling complexity in the text or picture on pupils' realistic solution of P-items. *An International Journal of Experimental Educational Psychology*, 1-13. doi:10.1080/01443410.2016.1259461

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Rolling Meadows, IL: Riverside.

Xin, A., Lin, C., Zhang, L., Yan, R. (2007). The performance of Chinese primary school students on realistic arithmetic word problems. *Educational Psychology in Practice: Theory, Research, and Practice in Educational Psychology, 23*, 145-159. doi:10.1080/02667360701320853

Xin, A., & Zhang, L. (2009). Cognitive holding power, fluid intelligence, and mathematical achievement as predictors of children's realistic problem solving. *Learning and Individual Differences, 19*, 124-129. doi:10.1016/j.lindif.2008.05.006

Yoshida, H., Verschaffel, L., & De Corte, E. (1997). Realistic considerations in solving problematic word problems: Do Japanese and Belgian children have the same difficulties? *Learning and Instruction, 7*, 329-338.

Zwann, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.

Chapter 3

Abstract

Realistic word problems are mathematical word problems that require the use of real-world knowledge to solve them. Research investigating children's ability to use realistic information in these problems has largely focused on interventions aimed at increasing students' realistic responses. The present study consists of three experiments that also had the same goal. In Experiment 1, only boys showed an increase in realistic responses when asked to produce a response sentence, and only on the division-with-remainder word problems. In Experiment 2, students showed a significant decrease in realistic responses when they were shown an example of how to answer a realistic word problem. In Experiment 3, the goal was to replicate the work of Palm (2008) who was able to triple the number of realistic responses by creating problems with more detail. Our sample of sixth-grade students failed to show an increase in realistic responses to the enhanced versions of realistic word problems. The results of these series of experiments indicate that a variety of intervention methods mostly prove to be insufficient to counter students already strongly held notions of how word problems are to be approached and answered.

Keywords: word problems, intervention methods, educational psychology, math cognition

Mathematical word problems have been a staple in math learning stemming back to the mid-fifteenth century because math word problems were of practical use for preparing people to participate in a world of trade, science, and engineering (Gerofsky, 2010). Today, math word problems remain to be one of the most common types of problems found in formal education (Jonassen, 2003). Math word problems (WPs) are designed and used for two reasons: (1) to practice using the basic arithmetic operations; and, (2) to train students to apply formal mathematical knowledge to real-world situations (Gravemeijer, 1997; Greer, Verschaffel, & De Corte, 2003; Verschaffel, Greer, & De Corte, 2000). While WPs have been an effective way of allowing students to practice using the basic arithmetic operations, they are still considerably more challenging than completing a basic math equation (Bjork & Bowyer-Crane, 2013). The opportunity of training students to apply formal math knowledge to math word problems that consider the real-world situation has been largely underutilized. Instead, students have developed a skill that has been termed, the 'suspension of sense making' (Schoenfeld, 1993), where the use of real-world information, when it comes to math in the math classroom, is not warranted and at times is even frowned upon (De Corte, Verschaffel, & Lasure, 1995; Verschaffel & De Corte, 1997a). As such, students find it very difficult to solve what are called "realistic" WPs – math WPs that require the consideration of real-world knowledge to be able to solve them.

Research examining student's performance on realistic WPs (i.e., the second function of math word problems) began with examining how students perform on these types of problems in general (Verschaffel, De Corte, & Lasure, 1994) and then moved onto ways to improve student's performance on these problems (Dewolf, VanDooren, &

Verschaffel, 2011; Palm, 2008; Silver, Shapiro, & Deutsch, 1993; Yoshida, Verschaffel, & De Corte, 1997). Early work in this line of research sought to reveal how students perform on word problems that require the problem solver to consider how their solution is impacted, if it is at all, by considering the real-world implications in the problem text (Greer, 1993; Verschaffel et al., 1994). Realistic WPs are those that, if taken under consideration of the typical conditions for that situation outside of school, the straightforward application of an arithmetic operation may not be appropriate (Palm, 2008). For example, take this WP: "John's best time to run 100m is 17 sec. How long will it take John to run 1 km?". If one were to seriously consider the realistic aspects of this problem, then the WP cannot be solved by simply cross-multiplying (e.g., 100m/17 sec = 1000m/170sec), because John's pace is going to slow if he is running a distance further than 100m; we also cannot forget that the problem text describes John's *best* time, and that John is unlikely to run that fast all of the time. However, these types of considerations are not considered in elementary school classrooms and children perform quite poorly on these types of WPs.

This paper describes three experiments that investigate manipulations expected to improve realistic responses (RRs) from elementary school students. As described in more detail below, previous research in this area has demonstrated that school children are resistant to including real-world knowledge when solving math WPs (e.g., De Corte et al., 1995; Greer, 1993; Silver et al., 1993; Verschaffel et al., 1994). Furthermore, previous attempts to increase realistic responding has had very little success (e.g., Dewolf et al., 2011; Reusser & Stebler, 1997; Xin, Lin, Zhang, & Yan, 2007; Yoshida et al., 1997). Given the importance of learning to apply real world information in problem solving situations, the manipulations employed in the proceeding studies provide opportunity to investigate ways to improve realistic responding.

Suspension of Sense-Making

Verschaffel and his colleagues (1994, 1995) and Greer (1993) were some of the first researchers to look at how children perform on realistic WPs. In these studies, the authors examined fifth- and sixth-grade students' use of realistic knowledge when solving math WPs that required the consideration and application of one's real-world knowledge. Students were given 10 WPs, five standard items, and five realistic items. Participants were asked to solve all the problems independently and they were not given assistance by the researchers or by their teachers. The test booklets provided a space for students to provide their numerical answers and a comments section. The comments section was designed to give students the opportunity to explain how they arrived at their answer, or any difficulties they encountered in the solution process for the problem.

Student responses were considered realistic (RR) if their numerical answer or the comments provided gave some indication that they had considered the real-world aspects of the WP. Students performed very poorly on the realistic items with only 17% of responses demonstrating the student's consideration of real-world knowledge, compared to 84% accuracy on the standard WPs (Verschaffel et al., 1994). In a follow-up study, with fifth-grade students, they replicated their earlier findings, showing 16% of responses were considered realistic (De Corte et al., 1995). More important, De Corte and colleagues were able to show that, despite not using their real-world knowledge, students understood that using their real-world knowledge would impact the numerical responses. With a certain degree of scaffolding – that is, students were shown a confederate peer response that was realistic and asked which was better, followed by a stronger scaffold where they were asked to draw a diagram of the problem situation – realistic responses could be increased. The number of RRs students gave could be improved by almost three times their original findings (e.g., 16% vs. 57%).

The results from these two early studies indicate that students are reluctant to include their real-world knowledge into the solution process of a math WP (De Corte et al., 1995; Verschaffel et al., 1994). However, they also illustrate that student's reluctance is not due to limited knowledge, as evidenced in the post-WP interviews (De Corte et al., 1995), but instead perhaps their resistance is a consequence of the culture created in a math classroom. Given this is so, research in the field has sought to find ways to increase RRs from elementary school students to these realistic WPs.

Interventions on the Suspension of Sense-Making

One set of studies has used warnings to increase RRs (Reusser & Stebler, 1997; Xin et al., 2007; Yoshida et al., 1997). With a group of fifth-grade students, Yoshida et al. (1997) randomly assigned Japanese students to a neutral or warning condition. Those in the warning condition saw a set of written instructions at the top of the sheet before the WP that told them:

> The test contains several problems that are difficult or impossible to solve because of certain unclarities or complexities in the problem statement. When you encounter such a problem, please write it down and explain why you think that you are not able to solve the problem. (p. 333)

They found no difference between the conditions in the number of RRs provided (15% vs. 20%, respectively) and their sample did not differ in the number of RRs compared to those originally reported by Verschaffel et al. (1994; 15% vs. 17%, respectively).

Reusser and Stebler (1997, Study 2) used a slightly older group of Swiss students (seventh-graders) and presented them with realistic WPs under a number of warning conditions (e.g., "Think about it carefully before you answer" or "Be careful! Some of the following problems aren't as easy as they look. There are, in fact, some problems in the booklet where it is very questionable if they are solvable at all", p. 319). They too found that the general warnings did not improve the RRs. However, when the warning was embedded in the WP itself, RRs increased to 42.5%. Xin et al. (2007, Study 2) also used different kinds of warnings with fourth- to sixth-grade Chinese students in the hopes of increasing RRs. These students were randomly assigned to a 'warning' condition or a 'process-oriented instruction' condition. In the 'warning' condition students were told that "some of the problems may not be as easy as they seem" (p. 151). In the 'process-oriented' condition students were asked to consider "…[T]he real-life situation behind the problem statement?", and "Is it appropriate to solve these problems by using straightforward arithmetic operations?" (p. 151). The instructions for the condition were printed on the front page of each respective booklet. The authors found a marginally significant difference between the groups (21.43% vs. 28.33%, respectively, $p = .055$), with students in the process-oriented instruction condition reporting more realistic considerations, e.g., responses in the comments area, but no differences in the number of 'correct' answers provided (16.7% vs. 19.67%, respectively).

Instead of using warnings, a set of researchers have investigated whether illustrations can prompt students to consider the real-world aspects of WPs (e.g., Dewolf, Van Dooren, Ev Cimen, & Verschaffel, 2014; Dewolf, Van Dooren, & Verschaffel, 2016; Weyns, Van Dooren, Dewolf, & Verschaffel, 2016). Focusing on students in fifth-and sixth-grade, these authors included pictorial representations of the situations described in the WPs (e.g., “Steve has bought 4 planks of 2.5m each. How many planks of 1m can he get out of these planks?” Accompanied by a picture showing a basket for waste pieces). In some cases, illustrations were paired with verbal and written warnings such as, “the test may also involve some difficult and tricky problems that are not as simple and straightforward as they may first seem” (p. 111) and in other cases students saw only an illustration, only a warning, or neither (Dewolf et al., 2014). Neither the illustrations or warnings increased the number of correct responses and they had a lower number of RRs than has been previously reported in the literature [e.g., 17% (Verschaffel et al., 1994) vs. 12.6% and 11.9% (Dewolf et al., 2014)].

Using a similar methodology, Weyns and colleagues (2016) presented fifth- and sixth-grade students with realistic WPs coupled with an illustration depicting the situation described in the WP or another illustration that cued the student’s attention to a realistic component in the problems. For example, Weyns et al. used an adapted version of the school WP [previously described above in a study by Wyndhamn and Säljö, (1997)] that asked how far apart a sports centre and a station were from each other. In one condition, students’ attention was cued by road signs that were added to the picture, but were drawn in a way that it remained unclear in which direction the ‘sports centre’ and the ‘station’ were. They also had a text condition in which an additional sentence was embedded at the end of the WP. For the school WP, the additional sentence was, “*Be aware that you do not know in which direction the sports centre and the station are*” (see Table 1, p. 6). They found overall that students were using RRs 27.9% of the time. While this is higher than is typically reported (17%), it still indicates students are using their realistic knowledge less than two-thirds of the time. Weyns et al. found that when the warning sentence was embedded in the WP, the number of RRs showed an increase compared to when no additional text was supplied (35.7% vs. 20.2%, respectively). However, no difference was found in the number of RRs with the basic and adapted pictures.

Dewolf and colleagues (2016) utilized eye-tracking equipment while participants completed realistic WPs with illustrative pictures, but they found that students were not even looking at the picture. When the pictures were presented in a way that forced students to view them (i.e., WPs and illustrations were presented sequentially rather than simultaneously), the pictures still did not improve their realistic responding. Given these findings, it is not surprising that the illustrations were not effective in the work done by Weyns et al. (2016). The results of studies using illustrations (Dewolf et al., 2016; Weyns et al., 2016) suggest that students do not use the illustrations as a tool to assist them in their problem solving solution.

There is an overwhelming amount of evidence illustrating that elementary school students are largely resistant to incorporating real-world knowledge into their solution process when presented with math WPs that necessitate the consideration of the realistic aspects to be effectively solved (Dewolf et al., 2014, 2016; Reusser & Stebler, 1997; Verschaffel et al., 1994; Weyns et al., 2016; Yoshida et al., 1997; Xin et al., 2007), despite understanding the realistic aspects in the problem text (De Corte et al., 1995; Wyndhamn & Säljö, 1997). The few examples of success involve warnings that are embedded right into the problem itself (Reusser & Stebler 1997; Xin et al., 2007). Nevertheless, there have been some studies that have succeeded in increasing realistic responses.

DeFranco and Curcio (1993) were able to increase the number of RRs given by sixth-grade students to a Division-with-Remainder WP: "328 senior citizens are going on a trip. A bus can seat 40 people. How many buses are needed so that all the senior citizens can go on the trip", (p. 61). Division-with-Remainder problems are typically found to be among the easiest of realistic WPs, with success rates typically near 50% (Silver et al., 1993; Verschaffel et al., 1994). In the first part of the study, DeFranco and Curcio had students read the problem and they were shown a possible answer of 8r8 or 8.5, and the students were asked if they agreed or disagreed with this answer. They found that only 10% of students believed neither answer was correct and instead thought 9 was the appropriate answer. One month later, in the second part of the study the context of the WP was altered to reflect the number of students in the class being examined:

> FACTS:
> *Date of party*: Friday, April 15
> *Time*: 4:00 – 6:00 PM
> *Place*: Ricardo's Restaurant, Queens
> *Number of children attending the party*: 32
> PROBLEM:
> We need to transport the 32 children to the restaurant so we need transportation. We have to order minivans. Board of Education minivans seat 5 children. These minivans have 5 seats with seatbelts and are prohibited by law to seat more than 5 children. How many minivans do we need?
> Once you have determined how many minivans we need, call 998-2323 to place the order (p. 62).

This time around students were asked to solve the problem and then call the company to place an order for the minivans. The person on the other line receiving the order was a researcher involved in the study. In this second part of the study, student's RRs increased to 90% – an increase of 80% as all 10% of students from part one of the study who had provided a realistic response continued to do so. This finding is extremely reassuring. The physical act of completing the problem (e.g., actually ordering the number of minivans needed) made it almost impossible for students to ask for a fraction of a minivan. Typically, students do not perform the physical acts which are being described in math word problems. Instead, students apply some mathematical operation to the numbers and report an answer. The result of DeFranco and Curcio's study supports the notion that there are rules to be followed when performing math WPs in the math classroom (Verschaffel & De Corte, 1985, 1997a). When children were removed from the context of the math classroom, as they were in the scenario designed by DeFranco and Curcio, the assumptions that the problem require only obvious calculations indicated in the problem no longer applied. Under the new context, children were free to use their real-world knowledge to complete the problem.

Taking a different approach, Palm (2008) altered the authenticity of the WP to increase RRs, but did not require students to physically engage in the WPs. Instead, Palm created a richer backstory in the problem text to better simulate some of the aspects of the real-life task situation. For example, here is the plank problem used by Verschaffel et al. (1994): "Steve has bought 4 planks of 2.5m each. How many planks of 1m can he get out of these planks?" (p. 276). The more descriptive version of this problem, used by Palm (2008) is:

> You are building a cabin and as walls you want to use planks that are 1m long. You are at the moment short of thirteen 1-meter planks. A friend says that she has found 4 planks, each 2.5m long. You are wondering if that is enough to finish the walls. How many 1-meter planks can you saw out of the planks she found? (p. 66)

As can be seen from the example of an authentic problem, there is more descriptive information, such as the purpose of needing additional planks, and why the planks are required to be 1m. This technique resulted in fifth-grade students providing RRs to the realistic WPs 51% of the time, a significant increase from what has been reported in the literature for the basic realistic WPs (e.g., 17% by Verschaffel et al., 1994). Palm (2008) suggests the more authentic versions of the realistic WPs allowed the students to be more engaged in the task context and resulted in more willingness to disregard the implicit assumption that the problem solution involves only producing and solving an equation. An alternative explanation could be that the more authentic versions of the realistic WPs enhanced students' comprehension of the text resulted in an improved mental representation of the situation. It is the children's improved comprehension that encouraged them to consider the relevance of the realistic nuances described in the WP.

The Current Studies

Three different experiments are reported here, each of which attempts to increase the number of RRs from students. The first experiment asked students to respond to each word problem using a full sentence. The second experiment provided students with an example of two different word problems, one standard and one realistic WP, both of which could be answered using the same mathematical operation(s). Students were shown and given an explanation of why the numerical answer to the realistic WP did not make sense given the situation described in the word problem. In the third experiment, students were given realistic WPs that were enhanced – that is, each problem was provided with more detail. Students were also asked to complete three multiple-choice questions, all of which could be answered by finding the appropriate content in the word problem

There have been mixed results in the literature on whether boys and girls respond differently to realistic WPs. Boaler (1993, 1994) has indicated that a WPs context (i.e., a gender bias) can impact how boys and girls respond to the problem, and that girls can be negatively affected by stereotypical female contexts where they may be more likely to erroneously include their real-world knowledge when it is not warranted. However, minimal work has been done examining gender differences in realistic WPs. In the few studies that have tested for gender differences, the general finding is that gender differences are not consistently found (Dewolf et al., 2014; Palm & Nyström, 2009), but gender differences have been found for particular WPs. The WPs that indicate gender differences tend to be those which involve situations that are gendered (e.g., sawing planks of wood). What appears to matter is whether to context of the problem is stereotypically gendered (Csikos, 2003; Palm & Nyström, 2009). More specifically, boys were more likely to answer realistically to the planks problem (Csikos, 2003), and girls were more likely to respond realistically to a horseback riding problem (Palm & Nyström, 2009). Because of these results, all of the studies reported here did include gender as a factor in realistic responding.

Experiment 1

As previously noted, Verschaffel and his colleagues (1994) posited that their Flemish sample performed better on realistic WPs because they were taught to respond to problems using a response sentence. Others have speculated that making students respond in sentences may force them to consider the real-world implications of their answer (e.g. Silver et al, 1993), but this speculation has not been explicitly tested. To explicitly test whether using a response sentence when answering realistic WPs helped students respond more realistically, students were randomly assigned to a control or written response condition. Those in the control condition were given both standard and realistic WPs using the same method and items used by Verschaffel et al. (1994) and

Greer (1993). In the written response condition, students responded to the same problems as those in the control condition, except that these students were asked to report their numerical responses in the form of a full sentence. It was hypothesized that students in the written response condition would perform better on the realistic WPs than the controls and the performance of the control participants would be similar to that found in the literature (e.g., 16-17% RRs; Verschaffel et al., 1994; Yoshida et al., 1997). The control group were also used to replicate the original findings of Verschaffel and colleagues (1994) and to ensure that the sample of Atlantic Canadian elementary school students is comparable to the previous findings. It was also hypothesized that the producing a sentence would not affect the accuracy of standard items.

Method

Participants

Eighty-four sixth-grade students (41 boys and 43 girls; M_{Age} = 11.73 yrs, SD = .360) participated from six elementary schools in a mid-sized Canadian city (these are the same students sampled in Chapter 2). As these students are under age, participants had received consent from a parent or guardian (see Appendix A). One participant was removed from the data analyses completely as this student had extreme scores on the Raven's Matrices ($z < -3.00$), the Woodcock-Johnson III Tests of Achievement composite scores ($z < -3.00$), and had 0% accuracy on the standard word problems. One other student had an extreme score on the Raven's Matrices ($z < -3.00$) but did not have extreme scores on the other assessments. This student was eliminated from all analyses involving the Raven's Matrices scores using a case-wise selection method in SPSS.

Eighty-three students were included in the following analyses (41 boys and 42 girls; M_{Age} = 11.73 yrs, *SD* = .363).

Materials

Raven's Standard Progressive Matrices (Raven, Raven, & Court, 1998). The Raven's Matrices is a standardized non-verbal, multiple-choice reasoning task used to assess problem solving ability, general intelligence, and cognitive ability (fluid intelligence). Raven and colleagues report a split-half internal consistency modal value of .91. The Raven's was designed to be useful for individuals of all ages and is made up of 60 problems – five series of 12 problems each – of diagrammatic puzzles with a part missing, which the test taker must find among the options provided. The initial problems are easy but the problems increase in difficulty as the child progresses through the test. Participants received a smaller selection of 28 items chosen based on age norms deemed appropriate for grade 6 children. Age norms range from children five and a half years-of-age to adults 85 years-of-age, which make it appropriate for the children to be sampled in this study. The Raven's is a group administered assessment.

Woodcock-Johnson III Tests of Achievement (WJ III Ach; Woodcock, McGrew, & Mather, 2001). The WJ III Ach is a multidimensional test of general intelligence for ages 2-90 years. The measure includes 22 tests for measuring skills in reading, mathematics, and writing, as well as important oral language abilities and academic knowledge, each of which measure one or more cognitive processes. Subsections of this measure were used for the current study, specifically those measuring reading comprehension and mathematical knowledge and ability. This standardized measure has undergone rigorous testing to ensure reliability and validity. All substests were administered one-on-one.

Calculation. Calculation is a test of math achievement measuring the ability to perform mathematical computations such as addition, subtraction, multiplication, division, and combinations of these basic operations, as well as some geometric, trigonometric, logarithmic, and calculus operations. The starting point was determined based on an estimate of the participant's present level of computation skill. The estimate is determined by the test makers and the starting point for grade 6 students is item 9. Calculation has a median reliability of .85 in the range of 5-19 year olds.

Math Fluency. Math Fluency measured accuracy on simple addition, subtraction, and multiplication facts. Participants were presented with a series of problems and were asked to complete as many as possible in the 3-min time limit. Math Fluency has a median reliability of .89 in the range of 5-19 year olds.

Passage Comprehension. Passage Comprehension began by examining symbolic learning. For example, the student would see an image of a house with a stick figure beside it and a statement with a word missing: "The house is _____ than the man". The student had to complete the statement so it described the image. Participants eventually moved onto reading short passages and identifying missing key words that made sense in the context of the passage. The items increased in difficulty by the removal of the pictorial stimuli and by increasing passage length, level of vocabulary, and complexity of syntactic and semantic cues. Passage Comprehension has a median reliability of .83 in the range of 5-19 year olds.

Applied Problems. Applied Problems required participants to analyze (e.g., determine the necessary math operation) and solve math problems. Participants must listen to the problem, recognize the procedure to be followed and then perform simple calculations. Each problem was visible to students but the experimenter read the problem aloud. As items can include extraneous information the person must identify not only the operation necessary to solve the problem but also the relevant information. Items increased in difficulty with more complex calculations. Applied Problems has a median reliability of .92 in the range of 5-19 year olds.

Reading Vocabulary. Reading Vocabulary measured skill in reading words and providing appropriate meanings. This test included three subtests: Synonyms, Antonyms, and Analogies. Synonyms required reading a word and providing a word that means the same, Antonyms required reading a word and providing a word that means the opposite, and Analogies required reading three words of an analogy and providing the fourth word to complete the analogy. Items increased in difficulty for each subtest. Reading Vocabulary has a median reliability of .87 in the range of 5-19 year olds.

Quantitative Concepts. Quantitative Concepts measured knowledge of mathematical concepts, symbols, and vocabulary. This test included two subtests: Concepts and Number Series. Concepts began with requiring the person to count and identify numbers, shapes, sequences, and further items required knowledge of mathematical terms and formulas. Number Series required the person to look at a series of numbers, figure out the pattern, and then provide the missing number in the series. Quantitative Concepts has a median reliability of .90 in the range of 5-19 year olds.

Realistic and Standard Mathematical Word Problems (Greer, 1993; Verschaffel et al., 1994). Realistic (R-items) and standard (S-items) WPs were taken from Greer (1993) and Verschaffel et al. (1994) as these problems have been used throughout the literature. Students received 10 WPs in total, five standard (S-items) and five realistic (R-items) problems. There were 10 possible S-items that could be included, and every S-item had a complimentary R-item. Four versions of the problem booklets were administered. One booklet was created with a random selection of five S-items and five R-items. Another booklet consisted of the five S- and five R-items that were not used in booklet one. The other two booklets comprised the same items as the first two booklets described but the WPs were presented in different orders. See Appendix D for list of problems.

Procedure

Schools were randomly assigned to the control or written response conditions rather than randomly assigning individual students. Randomization was done by school because each condition received a different set of instructions that were group administered. If random assignment were based on the individual, students in one condition would have to be absent from class when students in the other condition were not and vice versa. Random assignment by individual would have meant students were missing instructional time when their peers in the other condition were not. Random assignment by individual would also require the school to provide us with an additional room to test the two conditions. To reduce the burden placed on teachers and the institution, random assignment by school was viewed as the best option. Data collection occurred across multiple sessions. In the first session, the Raven's Matrices (Raven et al.,

1998) and Word Problem (WP) booklets were group administered. Students within each condition were randomly assigned to one of four versions of the WP booklet. In booklet version 1A the first question was the first WP on a list of 10 S-items, while the second problem was the second WP on a list of 10 R-items; in booklet version 2B all of the same WPs were used but the order changed so that the first question was the second WP on a list of 10 R-items and the second question was the first WP on the list of 10 S-items. In booklet version 1B the first question was the first WP on a list of 10 R-items and the second questions was the second WP on a list of 10 S-items; in booklet version 2A all of the same WPs were used by the order changed so the first question was the second WP on a list of 10 S-items and the second question was the first item on a list of 10 R-items. See Appendix D for the list of problems

Students in both the control and written response conditions were given the same, oral group instructions for the Raven's Matrices (Raven et al., 1998). The researcher illustrated two practice items with the group, explaining that the students had to fill in the missing piece of the picture with one of the options shown on the page and to write their response on the response sheet provided. Students had 25 minutes to complete the questions and the books were taken at the end of the allotted time. Students were then given instructions for the WP booklet. Students were asked to use the space on each page to show their work and to write their answer. Students were told that the researcher and their teacher could not help with the problems, and that if they are experiencing difficulties they should explain their problem in the comments section. The only difference in the instructions for the Word Problem booklet between conditions was that the students in the experimental condition were asked to report their answer using a full sentence. The additional time required to write the sentences in the experimental condition was estimated to take an additional 5 to 6 minutes.

The next visit was to conduct the one-on-one final interview. This interview was based on the protocol implemented by Palm (2008, Palm & Nyström, 2009) and was audio recorded. The final visits were individual testing sessions using the WJ III Ach measures for reading comprehension, mathematical knowledge, and achievement. Students were given the subtest in this order, Calculation Fluency, Math Fluency, Passage Comprehension, Applied Problems, Reading Vocabulary, and Quantitative Concepts. Three students were tested on the WJ III Ach in full and it was found that completing all six subtests at once proved to be time consuming and hard on the students, so the testing was divided into two phases, order 1: Calculation, Passage Comprehension, and Applied Problems, and order 2: Math Fluency, Reading Vocabulary, and Quantitative Concepts. The remaining students were then randomly assigned to be tested using order 1 or order 2 first and these orders were counter-balanced. As the WJ III Ach measure is standardized, test administration followed the existing protocol. Each testing session lasted 30-40 mins.

Coding

Verschaffel et al.'s (1994) coding scheme was used to code the responses. The coding scheme identified two kinds of information: (1) the numerical answer given in the 'answer area', and (2) written information in the 'comments area'. For this study, the focus was on the number of realistic responses (RRs). RRs were considered to be any response that had a realistic numerical answer or a non-realistic numerical answer that included a student's realistic consideration as written in the comments area. To obtain a measure of interrater reliability, 20% of booklets were coded by an independent rater using the same coding scheme. A high level of inter-rater agreement was established, κ = .828 (Cohen, 1960).

Accuracy on the standard-items was the sum of correct responses to any S-item with one point for each correct numerical response. A correct answer was considered the numerical answer found based on a straightforward calculation (e.g., EA; expected answer as outlined as part of the coding scheme). The maximum score for the S-items was five. Individual subtest and composite scores on the WJ III Ach were calculated using the computer software Compuscore and Profiles Program (Version 3.0, Schrank, & Woodcock, 2007). The current data was based on three composite scores from the WJ III Ach. Math Calculation Skills was the composite score of Math Calculation and Math Fluency, Math Reasoning was the composite score of Applied Problems and Quantitative Concepts, and Reading Comprehension was the composite score of Reading Vocabulary and Passage Comprehension). In Experiment 1, the standard scores based on age norms and raw scores were converted into standard scores with a mean of 100 and a standard deviation of 15.

Results and Discussion

The first step was to verify that the current study's sample replicated earlier findings (e.g., 16-17%, Verschaffel et al., 2000). The literature indicates that elementary school students provide realistic responses to word problems that require the use of real-world knowledge at a rate of 16-17%. In the current data, the realistic response (RR) rate was 21.46%, 95% CI [15.25%, 27.68%], and the response rate in the literature falls within this confidence interval. The percentage of participants producing RRs is therefore comparable to that found in previous research.

A 2 (Condition [Control, Written Response]) x 2 (Gender) ANCOVA with the three composite measures as dependent variables and controlling for scores on the Raven's Matrices was used to examine possible initial group differences. There was a significant main effect of Gender for the Math Calculation and Math Reasoning composite measures of the WJ III Ach, [$F(1, 67) = 16.400$, $p < .0005$, $\eta_p^2 = .197$ and $F(1, 67) = 4.781$, $p = .032$, $\eta_p^2 = .067$, respectively] such that boys scored significantly higher than girls. There was also a significant Condition by Gender interaction for the Math Reasoning composite measure, $F(1, 67) = 7.305$, $p = .009$, $\eta_p^2 = .098$. There was a significant difference in Math Reasoning scores between boys and girls in the written response condition, $F(1, 67) = 14.341$, $p < .0005$, $\eta_p^2 = .176$, where boys scored significantly higher than girls. There was also a significant difference between the conditions, but only for boys, $F(1, 67) = 6.375$, $p = .014$, $\eta_p^2 = .087$, where boys in the written response condition scored higher on Math Reasoning than boys in the control condition (see Table 1 for means). There was no main effect of Condition on the Raven's or on the WJ III Ach composite measures. Considering these findings, all remaining analyses examining Condition controlled for Math Reasoning scores.

It was hypothesized that there would be more RRs for those in the written response condition than in the control condition. This hypothesis was examined using a 2 (Condition [Control, Written Response]) x 2 (Gender) ANCOVA with Math Reasoning as the covariate. The current data did not support this hypothesis, as there was a no effect for Condition, Gender, or a Condition by Gender interaction [$F(1, 67) = .240$, $p = .626$,

Table 1
Estimated Marginal Means (SE) on WJ III Ach Composite Measures across Condition and Gender, while controlling for Raven's Matrices Scores

Measure	Condition			
	Control		Written Response	
	Boy	Girl	Boy	Girl
Reading Comprehension	92.661 (2.870)	94.487 (2.066)	95.630 (2.066)	94.594 (2.170)
Math Reasoning	98.600 (2.848)	99.816 (2.051)	107.483 (2.051)	96.217 (2.154)
Math Calculation	93.463 (3.340)	84.341 (2.405)	91.103 (2.405)	78.398 (2.526)

$F(1, 67) = .626$, $p = .431$, and $F(1, 67) = 2.648$, $p = .108$, respectively] (see Table 2 for means).

Although there was no overall main effect of requiring participants to write a sentence stating their conclusion to each problem, it is possible that the experimental manipulation might have increased RRs for division-with-remainder (DWR) problems. The two R-items (see Table 2, Buses and Balloon) that were DWR WPs were grouped and a 2 (Condition [Control, Written Response]) x 2 (Gender) ANCOVA, again controlling for Math Reasoning, was used to compare the average number of RRs for these two DWR problems between conditions. As before, there was no main effect of Condition or Gender [$F(1, 67) = 1.452$, $p = .233$, and $F(1, 67) = .001$, $p = .973$, respectively], but there was a significant Condition by Gender interaction, $F(1, 67) = 4.174$, $p = .045$, $\eta_p^2 = .059$. Follow-up analysis revealed that boys provided significantly more RRs to DWR WPs in the experimental condition, while there was no difference between conditions for girls [$F(1, 67) = 4.446$, $p = .039$, $\eta_p^2 = .062$ and $F(1, 67) = .521$, $p = .473$, $\eta_p^2 = .008$, respectively] (see Figure 1).

Table 2

Mean Percentages for Realistic Responses (RRs) to Realistic-items and Mean Percent Accuracy for Standard-items across Condition and Gender (N = 82)

Word Problems	Condition			
	Control		Written Response	
	Boys	Girls	Boys	Girls
Realistic-items				
Birthday Party	0	0	0	0
Temperature	10	13	19	0
Planks	0	0	14	10
Busses	15	35	38	20
Running	0	0	0	0
School	5	0	5	0
Balloons	30	30	52	35
Age	0	4	0	0
Rope	10	0	0	0
ChristmasCard	15	44	24	10
Total	17 (n = 19)	25 (n = 22)	31 (n = 21)	15 (n = 20)
	21 (n = 41)		23 (n = 41)	
Standard-items				
Birthday Party	100	100	100	100
Shop-keeper	100	100	91	90
Planks	50	70	89	90
Toy car	14	25	25	0
Boat	86	75	88	60
Walking Tour	82	89	100	100
Candy Bar	50	58	56	22
Piggy Bank	40	67	60	63
Clothesline	63	55	86	17
BirthdayCard	22	68	90	78
Total	56 (n = 19)	66 (n = 22)	71 (n = 21)	56 (n = 20)
	62 (n = 41)		63 (n = 41)	

Accuracy was examined for the S-items using a 2 (Condition [Control, Written Response]) x 2 (Gender) ANCOVA with Math Reasoning as a covariate. Students scored 63% overall on the S-items, but there was no effect for Condition, Gender, or a Condition by Gender interaction [$F(1, 67) = 1.285$, $p = .261$, $F(1, 67) = .166$, $p = .685$, and $F(1, 67)$

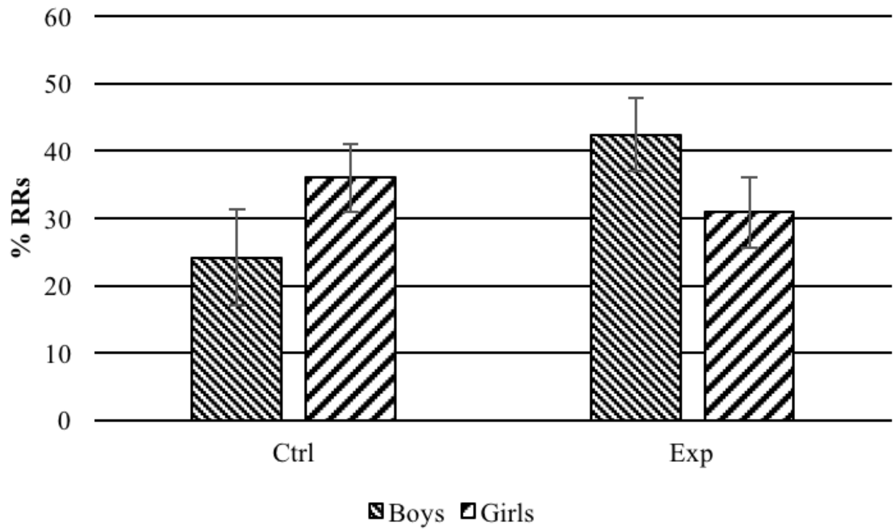

Figure 1. Estimated marginal mean percentage of realistic responses on Division-with-remainder (DWR) word problems by boys and girls across conditions while controlling for Math Reasoning composite measure of Woodcock-Johnson Tests of Academic Achievement III. All error bars are mean standard errors.

= .013, p = .910, respectively] (see Table 2 for means). It was hypothesized that the response sentence would not affect accuracy on S-items. Hypothesis two was supported as the additional prompt of asking students to report their answers in a full sentence had no impact on their standard WP solving performance. It should also be noted that the accuracy on S-items in the current sample (63%) was lower than what has been reported in the literature (i.e., 84%, Verschaffel et al., 1994).

The results of Experiment 1 confirmed that sixth-grade students in Atlantic Canada performed just as poorly as those in European and Asian countries (Verschaffel et al., 2000) on math WPs that require the consideration of real-world knowledge. However, having students respond using a full sentence did not significantly increase the incorporation of real-world knowledge into their solution process once prior math reasoning skills were accounted for, except in the narrow case of boys with DWR problems. It may be that using a response sentence was too subtle and was therefore not salient enough to produce an increase in the number of realistic responses. In Experiment 2, another approach was taken that would be more salient to students. The current experiment aimed to increase realistic responding by including an example of how to answer a realistic WP (see Figure 2).

Experiment 2

Previous research has indicated that explicitly warning students of possible difficulties or impossibilities associated with a particular problem still did not increase the number of realistic responses from those who received neutral instructions (Yoshida et al., 1997). Older students have been shown to increase their RRs but only when a warning was embedded into each problem (Reusser & Stebler, 1997). Fourth- and fifth-graders were explicitly asked to consider the real-life situations and to determine whether straightforward arithmetic operations are appropriate for solving these problems (Xin et al., 2007). Those asked to consider the real-life situation had a marginal increase (p = .055) in the number of realistic responses.

While there have been some improvements in the number of realistic responses given by students based on warning instructions (Reusser & Stebler, 1997; Xin et al., 2007), the number of RRs are still alarmingly low. Experiment 2 provided students with an example item to show them the difference between a standard and realistic WP. Although studies that have included warning instructions are prompting students on the difficulties of such problems, they are not equipping students with the expertise to solve the problems effectively. As part of Experiment 2, along with one appropriate answer for such a question, students saw an example of a standard word problem and an analogous word problem that required the consideration of real-world knowledge. Providing an

Example 1

A girl is counting drops from her leaky faucet. In one minute she counts 9 drops. How many will she count in the next 3 minutes? Write your answer in a complete sentence.

1 minute = 9 drops
1 minute = 9 drops
<u>1 minute = 9 drops</u>
3 minutes = 27 drops

Answer: She will count 27 drops in the next 3 minutes.

Comments:

Example 2

A girl is writing down names of animals that begin with the letter C. In one minute she writes down 9 names. How many will she write in the next 3 minutes? Write your answer in a complete sentence.

1 minute = probably less than 9
1 minute = a lot less than 9
<u>1 minute = maybe 3</u>
3 minutes = less than 27, maybe 18

Answer: She will probably name, maybe 18 animal names beginning with C.

Comments: <u>She is going to run out of animal names that begin with the letter C. So each minute after she will think of fewer animal names than the minute before</u>

Figure 2. Examples provided to participants in the experimental condition in Experiment 2.

example was expected to indicate to students that not all math WPs can be solved simply by applying a known mathematical operation, that some problems require consideration of the real world, and that some problems may not have an answer. It was hypothesized that providing an example of a problem that required consideration of the problems constraints would increase in the number of RRs.

Method

Participants

A sample of 144 sixth-grade students (65 boys and 79 girls; M_{Age} = 11.27 yrs, *SD* = .515) was recruited from eight elementary schools in a mid-sized Atlantic Canadian city. Information on ethnicity was not collected; however, the city largely consists of a Caucasian population. The participating schools included those from low socioeconomic areas to those in the highest socioeconomic areas. Five participants who had extreme scores on the Raven's Standard Progressive Matrices measure ($z < -3.00$), were eliminated from analyses using a case-wise selection method. One hundred and thirty-nine students were used in the following analyses (62 boys and 77 girls; M_{Age} = 11.28 yrs, *SD* = .515).

Materials

Raven's Standard Progressive Matrices (Raven, Raven, & Court, 1998). See description from Experiment 1.

Realistic and Standard Mathematical Word Problems (Greer, 1993; Verschaffel et al., 1994). See description from Experiment 1.

Procedure

Rather than randomly assigning individual students to condition, schools were randomly assigned to the written response or examples condition. This method was used because it meant less disruption to class time and less time taken away from classroom activities. Data collection occurred in one visit. Students completed the Raven's Matrices (Raven et al., 1998) first and then the Word Problem (WP) booklets. Students within each condition were randomly assigned to one of four versions of the WP booklet (see Experiment 1 for description).

Students in both the written response and examples condition were given the same oral group instructions for the Raven's Matrices (Raven et al., 1998; see Experiment 1 for detailed instructions). Following completion of the Raven's Matrices, instructions for the WP Booklets were given. Students in the control condition received the same instructions as those in the written response condition of Experiment 1, while students in the example condition received a different set of instructions. Students in the example condition were shown two examples on the front page of the WP Booklet. They were shown a standard word problem, how to solve it, and where to place the answer. They were also shown an example item of a corresponding realistic word problem with various ways to solve the problem, how more than one answer could be attained, or that there may not be enough information to find an answer (see Figure 2).

Coding

The coding scheme was the same as it was in Experiment 1, for a detailed description see Coding section of Methodology in Experiment 1 or Verschaffel et al. (1994).

Results and Discussion

To ensure that the randomization process was successful, a 2 (Condition [Written Response, Examples]) x 2 (Gender) ANOVA was used to examine mean differences on the Raven's Standard Progressive Matrices scores. The main effect of Condition was significant, $F(1, 135) = 7.434$, $p = .007$, $\eta_p^2 = .052$, where students in the written response scored higher on the Raven's than those in the examples condition (see Table 3). There was no effect of Gender or a Condition by Gender interaction [$F(1, 135) = 1.582$, $p = .211$ and $F(1, 135) = 2.057$, $p = .154$, respectively]. For that reason, all subsequent analyses involving condition controlled for scores on the Raven's.

Table 3

Estimated Marginal Means (SE) on Raven's Standard Progressive Matrices across Condition and Gender (N = 139)

Measure	Condition			
	Written Response		Examples	
	Boys	Girls	Boys	Girls
Raven's Matrices	22.474 (.762) n = 19	22.346 (.652) n = 26	19.465 (.603) n = 43	21.412 (.553) n = 51
Total	22.410 (.596) n = 45		20.483 (.409) n = 94	

A 2 (Condition [Written Response, Examples]) x 2 (Gender) ANCOVA, was used to examine mean percent RRs given for the realistic WPs, while controlling for scores on the Raven's. The main effect of Condition was significant, $F(1, 134) = 4.817$, $p = .030$, $\eta_p^2 = .035$, where students in the control condition reported more RRs than those in the examples condition (see Figure 3). There was a marginal effect of Gender, $F(1, 134) = 3.496$, $p = .065$ $\eta_p^2 = .025$, with boys having more RRs than girls, and there was no Condition by Gender interaction, $F(1, 134) = 2.357$, $p = .127$. Based on these findings, there was no evidence to support our a priori hypothesis. It can be concluded that the

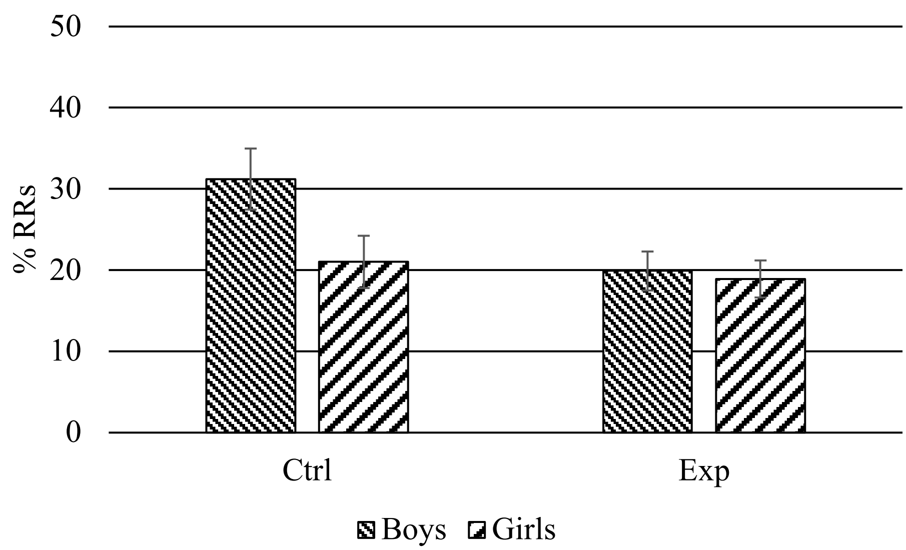

Figure 3. Estimated Marginal Mean Percentage on Realistic Responses to Realistic WPs across Condition, controlling for Raven's Standard Progressive Matrices Scores. All error bars are standard errors.

experimental manipulation did not improve the number of RRs provided to realistic WPs. If anything, it had a deleterious effect (see Table 4).

In Experiment 1, there was no difference in RRs between conditions on division-with-remainder (DWR) problems. Here in Experiment 2, DWR WPs were again evaluated on the number of RRs provided by students. A 2 (Condition [Written Response, Examples]) x 2 (Gender) ANCOVA, controlling for scores on the Raven's Matrices found no effect of Condition, Gender, or a Condition by Gender interaction [$F(1, 87) = .243$, $p = .623$, $F(1, 87) = .147$, $p = .703$, and $F(1, 87) = .143$, $p = .707$, respectively].

Students' accuracy on S-items were tested using a 2 (Condition [Written Response, Examples]) x 2 (Gender) ANCOVA, controlling for scores on the Raven's Matrices. On average students were getting 57% accuracy on the standard WPs. There

Table 4

Mean Percentages for Realistic Responses (RRs) to Each R-item across Condition and Gender (N = 139)

Word Problems	Condition			
	Control		Examples	
	Boys	Girls	Boys	Girls
Realistic-items				
Birthday Party	0	0	0	0
Temperature	42	8	12	12
Planks	5	12	9	8
Busses	53	19	26	16
Running	0	0	0	0
School	11	8	0	0
Balloons	26	50	28	37
Age	0	0	0	2
Rope	5	0	2	2
ChristmasCard	21	15	14	19
Total	33 (n = 19)	22 (n = 26)	18 (n = 43)	19 (n = 51)
	27 (n = 45)		19 (n = 94)	
Standard-items				
Birthday Party	100	100	100	100
Shop-keeper	88	83	82	90
Planks	73	80	79	83
Toy car	50	0	50	10
Boat	89	67	55	43
Walking Tour	75	87	90	83
Candy Bar	64	70	28	44
Piggy Bank	43	64	49	57
Clothesline	55	67	53	41
BirthdayCard	67	43	62	44
Total	65 (n = 19)	62 (n = 26)	55 (n = 43)	54 (n = 51)
	64 (n = 45)		57 (n = 94)	

was no effect of Condition, Gender, or a Condition by Gender interaction [$F(1, 134) = .963$, $p = .328$, $F(1, 134) = 1.804$, $p = .181$, and $F(1, 134) = .395$, $p = .531$, respectively].

The results indicate that providing examples of the way one could approach answering realistic problems did not have any beneficial effect on the number of RRs from students. In fact, the examples seemed to have a deleterious effect such that performance in the group given example solutions was worse than the students who did not receive the example solutions. In conclusion, the challenges students face with these kinds of WPs are not because they have not been shown how to solve problems, but that it is more likely that their prior experiences and the culture surrounding classroom mathematics has not facilitated this skill (see Verschaffel & De Corte, 1997a).

Experiment 3

Recent research using eye-tracking software have shown that students often are not even attending to supplementary information in WPs (e.g., illustrations, see Dewolf et al., 2015), and that students are often scanning the content to look for numerals and keywords that will assist them in the solution process (De Corte, Verschaffel, & Pauwels, 1990). Part of the aim of Experiment 3 was to ensure students are reading the problem fully before they attempt to solve the WP. In order to ensure that students will have read all parts of the problem at least once, a series of multiple choice questions that could be answered by reading the problem text were included in the WP booklets. The other goal of Experiment 3 was to replicate the findings of Palm (2008), who was able to triple the number of RRs by using WPs that had their authenticity enhanced by using a richer backstory with extra detail.

Method

Participants

Eighty-six sixth-grade students (47 boys, 39 girls, and 1 unidentified; M_{Age} = 11.89 yrs, *SD* = .285) were recruited from two elementary schools in a mid-sized Atlantic Canadian city. Information on ethnicity was not collected; however, the city largely consists of a Caucasian population. The participating schools included those from varying socioeconomic areas. Due to errors in the WP booklets, 20 students (14 boys, 6 girls) were removed from the data analysis. Students from the control group (e.g., Written Response) in Experiment 2 were also included in the current study as a control condition (n = 45). Two students were outliers on the Raven's Matrices ($z < -3.00$) and were eliminated from analyses using a case-wise selection method. The final sample included 111 students (52 boys and 58 girls, and 1 unidentified; M_{Age} = 11.747 yrs, *SD* = .343).

Materials

Raven's Standard Progressive Matrices (Raven, Raven, & Court, 1998). See description from Experiment 1.

Reading Comprehension-like Realistic and Standard Mathematical Word Problems (Palm, 2008; Verschaffel et al., 1994). Realistic and standard WPs were taken from Palm (2008) and Verschaffel et al. (1994). Each student received a pencil and paper task with six problems, three standard (S-) and three realistic (R-) items. There were six R-items and their corresponding six S-items. The problem text was at the top of the page, followed by three multiple-choice questions asking students for information found in the problem text. The fourth question asked students to complete the math word problem (see Appendix F). Only six problems were selected because of the extra time required for the multiple-choice questions. These six questions were a subset of the questions used in Experiments 1 and 2. Eliminating two difficult items made the test easier and reduced the time to complete it. Four versions of this task were used to ensure all S-items (six) and R-items (six) were seen by students, even though a student would have only answered three of the six S-times and three of the six R-items. Using four versions ensured that the WPs were not always seen in the same order. After the three multiple-choice questions on the content of the problem, the fourth question asked the student to solve the problem (in the form of a sentence) and included an area for comments. Students were not provided with assistance during this task and were asked to complete all problems on their own.

There were two conditions for the word problems: the realistic word problems developed by Greer (1993) and Verschaffel et al. (1994) called the Basic WPs, and the modified problems developed by Palm (2008) called the Authentic WPs. The modified items were considered to be more authentic in that they provided more detail and emphasized the realistic aspects. Two of the items (Planks and Balloons) were from Palm (2008) because these were enhanced versions of these problems while the other four were enhanced by the experimenters. Students received six authentic (three R-items and three S-items) or six of the basic realistic WPs (three R-items and three S-items). Instructions provided to students were the same for each condition.

Procedure

Students were given the Raven's Matrices task and were randomly assigned to the Basic (R-items) and Authentic (authentic R-items) conditions. All assessments were group administered. Because of the randomization of which items were R-items and which items were S-items, some of the participants in the control group had 4 R-items and 2 S-items, some had 2 R-items and 4 S-items, and some had 3 and 3. To adjust for this, the dependent variable is the analyses was the mean number of the R-items rather than the actual number of RRs and the mean number of S-items rather than number of items answered correctly.

Coding

The coding scheme was the same as it was in Experiment 1, for a detailed description see Coding section of Methodology in Experiment 1 or Verschaffel et al. (1994). As the control group was from Study 2, where they completed 10 problems, codes were only used for the same six problems that were used by the experimental group.

Results and Discussion

To ensure that the randomization process had not resulted in differences in Raven's scores across conditions, a 3 (Condition [Control, Basic, Authentic]) x 2 (Gender) ANOVA was used with scores on the Raven's Matrices as the outcome variable, and found no difference in Raven's scores between Conditions, Gender, or a Condition by Gender interaction [$F(2, 104) = .851$, $p = .430$, $F(1, 104) = .285$, $p = .595$, and $F(2, 104) = .203$, $p = .816$, respectively] (Table 5).

Table 5

Estimated Marginal Means (SE) on Raven's Standard Progressive Matrices across Condition and Gender (N = 110)

Measure	Condition					
	Control		Basic		Authentic	
	Boys	Girls	Boys	Girls	Boys	Girls
Raven's Matrices	22.474 (.762) n = 19	22.346 (.652) n = 26	23.167 (.744) n = 18	23.467 (.815) n = 15	22.133 (.815) n = 15	22.941 (.765) n = 17
Total	22.410 (.476) n = 45		23.317 (.552) n = 33		22.537 (.559) n = 32	

Before testing for differences between experimental groups, we wanted to ensure that the multiple-choice questions had the desired effect of making students read the problems and pay attention to the content. To do so, a 2 (Condition [Basic, Authentic]) x

2 (Gender) ANOVA was conducted with the mean number of correct responses on all multiple-choice across R-items, and S-items. There was no effect of Condition, Gender, or a Condition by Gender interaction (Table 6) and accuracy was high overall, $M = 86.5$.

Table 6

Estimated Marginal Mean Percent Accuracy on all Multiple-choice items and for R- and S-items across Condition and Gender (N = 65)

Problem Type	Condition			
	Basic		Authentic	
	Boys	Girls	Boys	Girls
Realistic-items	82 (n = 18)	83 (n = 15)	88 (n = 15)	90 (n = 17)
Total	83 (n = 33)		89 (n = 32)	
Standard-items	91 (n = 18)	87 (n = 15)	84 (n = 15)	88 (n = 17)
Total	89 (n = 33)		86 (n = 32)	
All items	87 (n = 18)	85 (n = 15)	86 (n = 15)	89 (n = 17)
Total	86 (n = 33)		87 (n = 32)	

To determine if students were providing more RRs to the R-items across conditions a 3 (Condition [Control, Basic, Authentic]) x 2 (Gender) ANOVA was conducted. No differences were found in the number of RRs between Condition, Gender, or a Condition by Gender interaction [$F(2, 104) = .210$, $p = .811$, $F(1, 104) = .349$, $p = .556$, and $F(2, 104) = 2.303$, $p = .105$, respectively] (see Table 7). There were no analyses examining DWR WPs as there was only one item in the current experiment that used DWR. Consistent with the results of Experiments 1 and 2, and the existing literature on DWR WPs (e.g., Silver et al., 1997; Verschaffel et al., 2000), the highest and most consistent level of RRs was to the Balloons WP. The one aberration from this pattern was the unusually high score on the Temperature WP for boys in the control condition (see Table 7). It is not clear why this particular subgroup did that well on that particular question.

Table 7

Mean Percentages for Realistic Responses (RRs) to Each R-item across Condition and Gender (N = 110)

Word Problems	Condition					
	Control		Basic		Authentic	
	Boys	Girls	Boys	Girls	Boys	Girls
Realistic-items						
Birthday Party	0	0	0	0	0	0
Temperature	42	8	11	13	13	6
Planks	5	12	11	7	13	24
School	11	8	6	0	0	6
Balloons	26	50	61	40	40	59
Rope	5	0	0	7	0	6
Total	36 (n=19)	23 (n=26)	29 (n=18)	22 (n=15)	22 (n=15)	33 (n=17)
	28 (n = 45)		26 (n = 33)		28 (n = 32)	
Standard-items						
Birthday Party	100	100	100	100	100	86
Shop-keeper	88	83	100	100	67	90
Planks	73	80	100	88	100	100
Walking Tour	75	87	92	100	100	80
Candy Bar	64	70	67	25	0	0
Clothesline	55	67	60	88	67	71
Total	76 (n=19)	88 (n=26)	95 (n= 18)	89 (n=15)	84 (n=15)	80 (n=17)
	83 (n =45)		92 (n = 33)		82 (n = 32)	

The students performed quite well overall on the S-items, averaging 86%, which is much higher than was reported in Experiments 1 and 2 (63% vs 57%, respectively). While there were no anticipated differences between the conditions on the accuracy of the S-items, a 3 (Condition [Control, Basic, Authentic]) x 2 (Gender) ANOVA with percent accuracy on S-items as the dependent variable was used to test this. There was no effect of Condition, Gender, or a Condition by Gender interaction [$F(2, 104) = 2.396$, $p = .096$, $F(1, 104) = .897$, $p = .962$, and $F(2, 104) = 2.225$, $p = .113$ respectively].

The results do not support those previously indicated by Palm (2008) in that authentic R-items did not facilitate RRs in the current sample. There was also no effect of having children read the problems three times and answering the multiple-choice items. The number of RRs provided by those in the Basic and Authentic WP conditions was not different than those provided in the Control condition. Furthermore, using multiple-choice items to prompt students on their realistic thinking did not improve RRs to realistic WPs.

Discussion

The aim of the current studies was to determine the effects of manipulating children's response method (Experiment 1), the instructional format (Experiment 2), and the word problem content (Experiment 3) with the goal of increasing the number of realistic responses (RRs) provided to realistic word problems (WPs). In each experiment, the manipulation was not effective and did not increase the number of RRs students gave when solving realistic WPs. In the case of Experiment 2, the manipulation had a deleterious effect and those in the example condition reported fewer RRs than those who simply wrote a response sentence. Many interventions in this line of research have been unsuccessful in eliciting an increased number of RRs (e.g., Dewolf et al., 2014; Reusser & Stebler, 1997; Yoshida et al., 1997; Xin et al., 2007), and, therefore finding that our manipulations did not provoke more RRs was not surprising. The lack of results supports previous conclusions about the difficulty of solving WPs realistically. Nevertheless, there are some patterns in the data from which we can take lessons.

One conclusion drawn is that students in this sample had the same relative success rate on realistic problems as Verschaffel and colleague's (1994) original reporting of elementary school student's poor performance on realistic WPs (e.g., 16-17%) has been consistently found by many others who have studied these types of WPs (Verschaffel et al., 2000). Not surprising, the students in these series of experiments continue to indicate that over the past couple decades, not much has changed. The number of RRs given by students, in the current sample, ranged between 21%-28%. The maximum average number of RRs was found in Experiment 3, but this number is likely inflated as students were given only six realistic items, and one item (that was eliminated) had a 0% RR rate in Experiments 1 and 2 (the Running WP). Furthermore, children in the control group of Experiment 3, who did poorly in Experiment 2, did just as well as the two experimental groups from Experiment 3 on this select group of items.

At first it may seem alarming that children have been ignoring, and continue to ignore, aspects of their real-world when solving math WPs. It is crucial to remember that the sixth-grade students that we sampled have an entire educational career of experience telling them their method is correct. Children have learned that the appropriate way to answer a math word problem is to apply a mathematical operation to the numbers in the problem and report that numerical answer. Furthermore, the math operation to be applied is the one the current math unit is covering. As such, students are active participants in a *didactic contract* – "a set of reciprocal expectations and obligations between the teacher and the students that has evolved in their ongoing interaction" (Brousseau, 1990 as cited in Gravemeijer, 1997) – and understand the expectations and responsibilities associated with completing WPs in the classroom setting (Brousseau, 1986). When it comes to mathematical modelling in the classroom, students are primarily exposed to standard WPs, to be solved using the straightforward application of one of the four basic arithmetic operations, with a limited amount of estimation (Reusser, 1988; Reusser & Stebler, 1997; Verschaffel et al., 2000). It is therefore quite reasonable for children to respond to realistic WPs in a non-realistic way and do only the obvious calculation indicated by the wording of the problem. In this view, WPs are generally considered artificial. The WP is simply a facade for practicing the arithmetic skills that had been recently covered (Maier, 1991 as cited in Palm, 2008). The children responding to these realistic WPs are doing exactly what they have been instructed to do for years.

Experiment 2 was an attempt to illustrate to students another way of thinking about and answering WPs, but it was not successful. It is possible that providing the example may have created self-doubt in those students. Students may have thought they understood the WP but after reviewing the example, they may have reconsidered their original thought process as flawed. Along that line of reasoning, the example may not have been helpful because, while the example demonstrated a specific situation, it did not readily map onto the situations described in the other WPs and did not facilitate realistic considerations. It may also have been difficult for students to flip back and forth between answering some questions realistically while other problems were standard problems. On the other hand, a simpler explanation is that one example was not enough to override years of training.

Despite the difficulty of this study and past studies in improving realistic responding, it is surprising that there was no effect for the authentic problems in Experiment 3. Palm (2008) has previously reported a substantial increase in RRs to realistic WPs when the authenticity has been enhanced (e.g., when a more detailed description of the situation was provided). However, participants were not any more successful at WPs with their authenticity enhanced than they were with the non-authentic items.

One difference between our version of the authentic problems and those of Palm was the addition of the multiple-choice questions. The addition of these questions could have distracted from the authentic cues created by the enhanced problems, even though they were designed to draw attention to these cues. Perhaps students were attending to the authentic cues during the first iteration of reading the WP, but when they were asked to complete the multiple-choice items that information began to deteriorate. When it came time to provide a numerical answer to the WP students may have felt that because they had read the problem many times to answer the multiple-choice items that they did not need to read it again. If the authentic cues had deteriorated over that period of time and students reasoned that the WP had been read a sufficient amount of times, those cues would no longer be beneficial to the solution process. Alternatively, perhaps the difference between our findings and that of Palm (2008) is due to the selection of items that we used. However, if you consider the two items used in the current study that were also used by Palm (2008), the Planks and Balloons WPs, Palm (2008) reported a RR rate of 32% and 75%, respectively, while students in our study had RR rates of 14% and 50%, respectively. This suggests that lower performance on the authentic realistic WPs, compared to Palm (2008), is not explained by the selection of items.

Another difference between the authentic problems in Palm (2008) and the ones presented was that our R-items were interspersed with S-items. Palm (2008) only had R-items, and instead had authentic and non-authentic items as a between-subject comparison, rather than the within-subject comparison in the present study. In fact, all the experiments presented here had both R-items and S-items in the WP measure. It is possible that it is easier to respond realistically when all the problems in a set are realistic. When R-items are mixed with S-items (i.e., there are some items that can be answered the conventional way) this may encourage children to answer conventionally to all of them. This is reminiscent of the risks that cognitive biases have on problem solving, namely the cognitive bias mental set (Luchins, 1942). The mental set is a strategy that is typically employed because it has proved to be successful in the past. Mental sets can be useful for working through straightforward problems but may not be useful for novel problems. If students are viewing the realistic problems as being a straightforward problem they can easily fall into the trap of employing a problem-solving strategy that had been useful in the past but is not appropriate for solving the current problem. This possibility is worthy of further testing.

In Experiment 3, the problems were structured, by including multiple choice questions, to make children read and understand the content of the WP. It was hoped that this would make children more likely to answer realistically. Accuracy of the multiple-choice items suggest that students were in fact reading the problems entirely, as it ranged from 84%-90%. Nevertheless, children in the basic and authentic group, both of whom had these multiple-choice questions, did not do better than the control group on realistic responding. Therefore, student difficulty with realistic WPs cannot be chalked up to them not reading and understanding the content of the problem. Even though previous evidence has suggested that students often do not read WPs and instead just search for keywords (Dewolf et al., 2014), children were still no better at answering realistically even when prompted with questions to pay attention to the entire text of the problem.

Perhaps students read the text for the multiple-choice questions and still fell into old habits of looking for keywords to decide what equations to use.

A consistent finding in the literature is that RRs are provided more often for WPs that are division-with-remainder (DWR) problems than the other types (e.g., De Corte et al., 1995; Silver et al., 1993; Verschaffel et al., 1994), and data from Experiments 1 and 2 were consistent with this observation. It was difficult to assess performance on DWR problems in comparison to other types in Experiment 3 given that there was only one realistic DWR item. While there was no main effect on condition for DWR problems in Experiment 1, there was an interaction effect of intervention with gender. Boys who were asked to respond to their answers in a complete sentence reported more RRs to DWR WPs than the boys in the control condition, while no difference was found between conditions for the girls. However, it is not clear why the response sentence would help boys more than girls. Cooper and Harries (2002) suggest that boys may be more likely than girls to show a willingness to 'break the rules' when solving a DWR WP. It is possible that boys may feel more comfortable providing an estimated realistic answer rather than the actual computation answer, however that still does not explain why the response sentence was more effective for boys. In any case, this result suggest that more research should be done to look at sex differences with realistic WPs.

Conclusion

Mathematical modelling is an extremely important skill for individuals in our society. With the rapid technological changes occurring every day, now more than ever we need adults with strong mathematical modelling skills. One of the biggest gripes heard from young people in math is, 'when am I ever going to use this stuff', and they are not wrong for asking. The current attitude of North American youth is that the math they learn in school is not applicable to their lives. However, this attitude can be changed. The school culture needs to allow for more opportunities to show how math is not ‘set in stone’, there is not always ‘one answer’, and that the mathematical concepts that have been discovered throughout the ages have been done, in part, because of the observations made in the real-world.

The results presented here suggest that getting students to consider the real-world implications of the problems they are solving is no easy task. It cannot be accomplished through the production of a sentence that might not make sense, nor through an example, nor through providing extra details and making sure students read them. Instead, the implication is that training children to think realistically might take larger changes to how kids learn to look at WPs, perhaps by not treating WPs as simply a way to practice recently learned math techniques. It may be that problems should always be structured to include some real-world knowledge. If we want students to think more realistically and be able to think of math as something that can apply to the real-world, simple measures may not be sufficient. Instead, we may have to change the didactic contract.

References

Bjork, I. A., & Bowyer-Crane, C. (2013). Cognitive skills used to solve mathematical word problems and numerical operations: A study of 6- to 7-year-old children. *European Journal of Psychological Education, 28*, 1345-1360. doi:10.1007/s10212-012-0169-7

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

De Corte, E., Verschaffel, L., & Lasure, S. (1995, April). *Word problems: Game or reality? Studies of children's beliefs about the role of real-world knowledge in mathematical modeling.* Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, US.

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.

DeFranco, T. C., & Curcio, F. R. (1993). A division problem with a remainder embedded across two contexts: Children's solutions in restrictive vs. real-world settings. *Focus on Learning Problems in Mathematics, 19*, 58-72.

Dewolf, T., Van Dooren, W., Ev Cimen, E., & Verschaffel, L. (2014). The impact of illustrations and warnings on solving mathematical word problems realistically. *The Journal of Experimental Education 82*, 103-120. doi:10.1080/00220973.2012.745468

Dewolf, T., Van Dooren, W., Hermens, F., & Verschaffel, L. (2015). Do students attend to representational illustrations of non-standard mathematical word problems, and, if so, how helpful are they? *Instructional Science, 43*, 147-171. doi:10.1007/s11251-014-9332-7

Dewolf, T., Van Dooren, W., & Verschaffel, L. (2011). Upper elementary school children's understanding and solution of a quantitative problem inside and outside the mathematics class. *Learning and Instruction, 21*, 770-780. doi:10.1016/j.learninstruc.2011.05.003

Dewolf, T., Van Dooren, W., & Verschaffel, L. (2016). Can visual aids in representational illustrations help pupils to solve mathematical word problems more realistically? *European Journal of Psychology of Education, 32*, 335-351. doi:10.1007/s10212-016-0308-7

Gerofsky, S. (2010). The impossibility of 'real-life' word problems (according to Bakhtin, Lacan, Zizek and Baudrillard). *Discourse: Studies in the Cultural Politics of Edcuation, 31*, 61-73. doi:10.1080/01596300903465427

Gravemeijer, K. (1997). Commentary solving word problems: A case of modeling? *Learning and Instruction, 7*, 389-397.

Greer, B. (1993). The mathematical modeling perspective on wor(l)d problems. *Journal of Mathematical Behavior, 12*, 239-250.

Greer, B., Verschaffel, L., & De Corte, E. (2003). "The answer is really 4.5": Beliefs about word problems. In G. Leder, E. Pehkonen, & G. Torner (Eds.), *Beliefs: A hidden variable in mathematics education* (pp. 271-292). Secaucus, NJ, US: Kluwer Academic Publishers. Retrieved from http://www.ebrary.com

Jonassen, D. H. (2003). Designing research-based instruction for story problems. *Educational Psychology Review, 15*, 267-296. doi:1040-726X/03/0900-0267/0

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs, 54*(6), i-95. http://dx.doi.org/10.1037/h0093502

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 7*, 37-58. doi:10.1007/s10649-007-9083-3

Palm, T., & Nyström, P. (2009). Gender aspects of sense making in word problem solving. *Journal of Mathematical Modelling and Application, 1*, 59-76.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section* Reusser, K., & Stebler, R. (1997). Every word problem has a solution – The social rationality of mathematical modeling in schools. *Learning and Instructions, 7*, 309-327.

Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce on formal and informal mathematics. In *Informal Reasoning and Education.* J. F. Voss, D. N.

Schrank, F. A., & Woodcock, R. A. (2007). WJ III Normative Update Compuscore and Profiles Program (Version 3.0) [Computer Software]. *Woodcock-Johnson III.* Rolling Meadows, IL: Riverside Publishing.

Silver, E. A., Shapiro, L. J., & Deutsch, A. (1993). Sense making and the solution of division problems involving remainders: An Examination of middle school students' solution processes and their interpretations of solutions. *Journal for Research in Mathematics Education, 24*, 117-135.

Sheie, A. H. (2016). *Research based best practices for teaching mathematics and improving math attitudes* (Honors Senior Theses). Retrieved from http://digitalcommons.wou.edu/honors_theses

Verschaffel, L., & De Corte, E. (1997a). Word problems: A vehicle for promoting authentic mathematical understanding and problems solving in the primary school? In *Learning and Teaching Mathematics.* T. Nunes, & P. Bryant (Eds.) (pp. 69-97). East Sussex, UK: Psychology Press Ltd.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4*, 273-294.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making Sense of Word Problems.* Lisse, The Netherlands: Swets & Zeitlinger.

Weyns, A., Van Dooren, W., Dewolf, T., & Verschaffel, L. (2016). The effect of emphasizing the realistic modelling complexity in the text or picture on pupils' realistic solutions of P-items. *Educational Psychology, 37*, 1173-1185. doi:10.1080/01443410.2016.1259461

Woodcock, R. A., McGrew, K. S., & Mather, N. (2001, 2007). *Woodcock-Johnson III Tests of Avhievement.* Rolling Meadows, IL: Riverside Publishing.

Wyndhamn, J., & Säljö, R. (1997). Word problems and mathematical reasoning – A study of children's mastery of reference and meaning in textual realities. *Learning and Instruction, 7*, 361-382.

Xin, A., Lin, C., Zhang, L., & Yan, R. (2007). The performance of Chinese primary school students on realistic arithmetic word problems. *Educational Psychology in Practice: Theory, Research and Practice in Educational Psychology, 23* 145-159. doi:10.1080/02667360701320853

Yoshida, H., Verschaffel, L., & De Corte, E. (1997). Realistic considerations in solving problematic word problems: Do Japanese and Belgian children have the same difficulties? *Learning and Instruction, 7*, 329-338.

Chapter 4

General Discussion

In this dissertation, a group of studies are reported in which the ability of elementary school children to take real-world information into account when responding to mathematical WPs was investigated. In Chapter 2, the goal was to examine individual differences associated with children's ability to provide a realistic response (RR) to a math word problem that required consideration of the problem constraints. Chapter 3 consisted of three experimental attempts to increase the number of RRs sixth-grade students provide to realistic word problems. Although the results of each of these studies has already been discussed in their respective chapters, some of the implications of the findings are further explored and additional analyses of the data (presented in Appendix G) are examined in this final chapter.

Individual Differences

Verschaffel and his collaborators admitted that, "[A] limitation of the research is that…relatively little attention has been paid to individual differences" (Verschaffel, Greer, & De Corte, 2000, p. 156). Largely, the research on individual difference in math word-problem has focused on performance for standard word problems (Bjork & Bowyer-Crane, 2013; Kail & Hall, 1999; Muth, 1984; Pimperton & Nation, 2010; Vilenius-Tuohimaa, Aunola, & Nurmi., 2008). However, standard word problems in elementary school usually do not necessitate realistic thinking and typically only require a one-step solution using one of the four basic arithmetic operations (e.g., addition, subtraction, multiplication, and division). Nevertheless, reading comprehension and prior mathematical ability were often implicated as being related to WP performance and both have been found to be important predictors in performance on basic WPs (Bjork & Bowyer-Crane, 2013; Fuchs et al., 2006; Muth, 1984).

Unlike the findings in the literature on standard word problems, there was no evidence of individual differences in the current sample on either reading comprehension or mathematics ability. The first hypothesis tested was that individual differences in skills such as prior math calculation ability and reading comprehension would be unique and positive predictors of realistic responding by students on WPs that require the use of realistic consideration. However, this hypothesis was not supported. The results of regression analysis revealed that none of the individual differences, Raven's Matrices, WJ III Ach composite measures, and gender, uniquely accounted for a significant amount of variance in RRs. Math calculation skills was not a unique predictor of RRs, and the amount of variance contributed to RRs by reading comprehension was significant only at the $> .10$ level. Given the extensive evidence supporting reading comprehension as an important factor for success on standard WPs (Bjork & Bowyer-Crane, 2013; Muth, 1984; Pimperton & Nation, 2010; Vilenius-Tuohimma et al., 2008) the failure to find that reading comprehension was a predictor of success on realistic WPs was quite surprising. It may be that reading comprehension was not a significant predictor of success on realistic WPs because of its shared variance with math reasoning skills as these two variables held a strong correlation.

One explanation of the lack of an association between reading comprehension and performance on realistic problems – not explored in Chapter 2 – is due to low scores on reading comprehension or, more specifically, their scores on passage comprehension (see Appendix G). The sixth-grade students were scoring below age norms on the WJ III Ach composite measure on reading comprehension which was being driven by the low performance on passage comprehension. It is possible that reading comprehension would have been found to be related to realistic responding if a substantial part of the sample were at least average on this ability. However, that was not the case with the current sample.

The findings in the current study do not corroborate those reported by Xin and Zhang (2009), who found math achievement to be a predictor of success on both standard and realistic WPs. Our data failed to indicate that math calculation ability predicted RRs and success on standard WPs. This was an unexpected finding given the evidence in the existing literature which does support math ability as predicting RRs and accuracy on standard WPs (Kail & Hall, 1999; Muth, 1984; Xin & Zhang, 2009). Additional analyses (Appendix G) revealed that the sample of sixth-grade students in Chapter 2 were also scoring below age norms on the WJ III Ach composite measure of mathematical calculation ability. The mathematical calculation skills composite measure is made up of math fluency and math calculation skills, both of which were found to be below age norms. It remains possible that math achievement is a predictor of success on realistic WPs, but the current sample of students' low performance precluded being able to detect such a finding.

While not discussed in Chapter 2, another unexpected finding in the current study was that math reasoning positively predicted performance on standard WPs. Recall that math reasoning is a composite measure taken from the WJ III Ach and consists of achievement on applied problems and quantitative concepts. When this relation is broken down by the subtests associated with math reasoning, the subtest of math reasoning that focuses on knowledge of mathematical concepts and symbols remains (quantitative concepts) predictive while the other subtest of math reasoning, focusing on analyzing and solving math problems (applied problems), is not predictive (see Appendix G). It is interesting that the subtests focusing on number series and understanding of math symbols (e.g., %, +, x, =, <, etc.) is what is driving the relation and not the subtest that has students solving word problems.

It could be that the relation is being driven by knowledge of concepts and symbols rather than performance on word problems because, in the end, students must be able to calculate the response in order to solve a word problem. Learning math in most classrooms necessitates focussing on calculation strategies (Schumacher & Fuchs, 2012), so students are already primed to provide a calculated answer above all. It is also possible that, while the standard word problems used in this research are comparable to what a student would typically encounter in the math classroom, it may be that having standard items mixed with realistic items altered how familiar students felt these problems were.

Inoue (2008) has argued that when there is a lack of familiarity with a problem situation, students are more likely to engage in a computational approach (i.e., applying some mathematical operation to the numbers in the problem) than to apply mathematical reasoning. Perhaps the quantitative concepts subscale of the math reasoning composite measure are better predictors than applied problems (the other subscale in math reasoning) because students rely on that knowledge when they are faced with problems they do not readily recognize. It has also been suggested that when students complete word problems, rather than using their reasoning skills, they often focus on number, symbols, and calculations (Walkington, Sherman, & Petrosino, 2012) and that is why concepts of numbers and symbols is predicting performance on standard WPs rather than performance on realistic word problems.

The key finding in Chapter 2 was RRs continued to predict variance on standard WP performance after all other types of individual factors were controlled. There is something about being more likely to incorporate real world knowledge into the solution process that makes accuracy in standard WPs more likely. The ability to construct a situation model as a strategy for solving WPs is beneficial for both realistic and standard WPs. Perhaps the students who performed well on realistic items did so because they constructed a mental model of the problem situation (Greer, 1993; Zwann & Radvanksy, 1997). It could be that these students who are able to use realistic considerations to better solve realistic WPs are also better equipped to know when and when not to use this style of reasoning. Another possibility is that children who develop a mental model (e.g., visualizing the situation) have included the situational constraints within the mental model; thus, they have a more accurate and complete model. The problem may be that the children who overlook the relevant constraints are not constructing a complete mental model of the problems situation, and instead just look for clues to the arithmetic operations to be used.

Cooper and Dunne (2000) found that children who speculated or questioned the specifics on standard WPs (e.g., asking if there were already apples in the larger container in the Apples items, see Appendix C), and then realized that the problem would be unsolvable, were the same students who reported the greatest number of RRs to the realistic items. The authors considered that there may be some sort of individual predisposition responsible for these variations. It is possible that the students who asked questions or speculated, were trying to develop a mental model which included potential constraints within the problem text.

The following paragraph will summarize the current findings on individual differences. The data illustrated some of the important components used to solve realistic WPs, but were unable to demonstrate the importance of reading comprehension, although this may have something to do with our entire sample performing very poorly in passage comprehension skills (see Appendix G). A novel finding in the study indicates that there is a benefit to having and using realistic knowledge when solving standard WPs. The standard WPs were comparable to mathematical textbook problems, which means that having realistic knowledge is related to solving the types of word problems that are presented throughout one's primary and secondary mathematical education.

The main finding of Chapter 2 is a major contribution to the existing literature. Prior to this study, no other research has directly examined the relation between students' performance on basic math word problems and math word problems that require realistic consideration. It appears that there has been an assumption among those involved in this line of research that because one of the purposes of word problems is to engage in mathematical modeling it must also be related to the other purpose of word problems, practicing the basic arithmetic operations (Gravemeijer, 1997; Greer et al., 2003; Verschaffel, Greer, & De Corte, 2000). This notion no longer needs to be an assumption as the evidence provided in Chapter 2 has shown that realistic reasoning is an important skill that contributes to students' success on word problems that are used in everyday mathematical classrooms.

Improving Realistic Responding

In Chapter 3 three different ways to improve realistic understanding were explored. The research to date has consistently shown that elementary school students perform poorly when faced with word problems that require them to use their real-world knowledge (for a review see Verschaffel et al., 2000). The data in the preceding experiments further support these findings. The elementary school students tested here in Atlantic Canada, specifically those in the control condition (Experiment 1), performed at a similar rate to participants in earlier studies. In all three experiments, students' realistic response (RR) rate for realistic word problems ranged from 21% to 28%. This finding was not very surprizing, and, in fact, there was no anticipation that this sample of students would perform differently than what has traditionally been found in other populations. The low RR rate provided by students in the current sample is likely a result of their educational experience and their participation in the didactic contract – the reciprocal expectations held between the teachers and students in the mathematical classroom (Brousseau, 1986) – and not a result of students having insufficient real-world knowledge to effectively answer the problems. Students have participated in a set of rules and expectations within the classroom that dictates their role as the solver of a math word problem (e.g., didactic contract). From this perspective, they are using the 'correct' method for solving realistic WPs.

The series of experiments described in Chapter 3 were aimed at increasing the number of RRs students provide to realistic WPs by manipulating the response method (Experiment 1), the instructional format (Experiment 2), and the word-problem content (Experiment3). In all three experiments, the manipulation was not effective in increasing the number of RRs. This finding is consistent with many of the unsuccessful interventions used in this line of research (e.g., Dewolf, Van Dooren, Ev Cimen, & Verschaffel, 2014; Reusser & Stebler, 1997; Yoshida et al., 1997; Xin, Lin, Zhang, & Yan, 2007).

Using a response sentence to increase RRs to realistic WPs was spurred by the work of Silver, Shapiro, and Deutsch (1993) and Verschaffel, De Corte, and Lasure (1994). Silver et al. (1993) explained that student's poor performance on Division-with-Remainder (DWR) WPs (that varied in remainder size) in their sample may have been due to the lack of written explanation required in schools. Meanwhile, Verschaffel and colleagues (1994) suggested that their sample were doing so well on DWR WPs because Flemish students are required to give a response sentence in math. Overall, the written response sentence was found to be ineffective in increasing the total number of RRs provided by students, however, there was an interaction related to DWR WPs. There was no expectation that boys and girls would respond differently, as Silver et al. (1993) and Verschaffel et al. (1994) had not reported any gender differences in their sample. At the same time, there have been some gender differences with realistic WPs when they involved a gender-stereotyped situation (see Cooper & Dunne, 2000; Boaler, 1993, 1994; Palm & Nyström, 2009; Zohar & Gershikov, 2008). It is possible that the effect is being driven by one of the DWR WPs having a stereotypical male context. The Bus WP is about transporting soldiers to a training site and soldiers are usually thought of as being male, or even that trucks, cars, and other vehicles (e.g., a bus) are stereotypical male interests. Another possibility is that the boys felt more comfortable deviating from the actual numerical answer and providing an estimated answer (see Cooper & Harries, 2002).

Reporting the numerical answer in a full sentence may be particularly useful for a specific kind of word problem [e.g., division-with-remainder (DWR)]. There may be something unique to seeing a fractional amount written and explained in a mathematical context, e.g., “Each grandchild will get 4.5 balloons” that alerts one to the realities of the numerical response. While the numerical answer reported in a full sentence to other WPs may not be as transparent, e.g., “There will be 936 Christmas cards sold in January, February, and March”. Moreover, the response sentence provided for another WP may have zero transparency, “It will take John 170 seconds to run 1 km”. Perhaps the use of the response sentence can only be expected to be helpful when the statement could create a highly salient indicator of the response not matching the situation, e.g., a fraction of a balloon. The manipulation may not have been as successful as anticipated because previous research has demonstrated that students are not used to using a response sentence in math (see Silver et al., 1993) and they did not see the relevance of using it. Even today, a review of the curriculum guide for Newfoundland finds no use of responses sentences for word problems (Newfoundland Labrador, Education and Early Childhood Development, 2015). The response sentence may have been more useful if students were told why there were being asked to use it. If they had known that the response sentence was designed to help them understand if their numerical response fit the situation, they may have paid more attention to what they were writing, rather than blindly writing their numerical response in words.

Warnings have been heavily used by researchers to increase realistic responses to no avail (see Dewolf et al., 2014; Reusser & Stebler, 1997; Weyns, Van Dooren, Dewolf, & Verschaffel, 2016; Yoshida et al., 1997). It is possible that warnings (e.g., oral and written warnings, and those embedded in the word problems) may not have worked because students were being warned against an issue they have had no exposure to (i.e., sometimes the numerical answer needs to be adjusted to fit the situation being described). How are students supposed to know what to look out for? Experiment 2 participants were shown another way of thinking about and answering realistic WPs, but it was not successful. Instead, the examples had the opposite effect, showing a significant decrease in RRs. It is possible that the example depicted a specific situation and students were unable to generalize it to other types of problems. Another possibility is that the examples provided caused students to second guess their solutions. However, the most likely explanation is that one example is not strong enough to counter the didactic contract (Brousseau, 1986).

Despite the many efforts of those to increase RRs with no success (e.g., Reusser & Stebler, 1997; Verschaffel et al., 2000; Yoshida et al., 1997; Xin et al., 2007), Palm (2008) has previously reported a sizable increase in RRs to realistic WPs when the WP was given with a more detailed description of the situation, Palm referred to the more detailed problems texts as having enhanced authenticity. The results of the current data contrast with those reported by Palm. The data failed to replicate Palm's finding as there was no increase of RRs in the current sample using similarly enhanced problems (and some problems that were exactly the same). There were, however, several methodological differences between Experiment 3 and Palm's (2008) works that may account for the differences in our findings. First, the current study included three multiple-choice items after each WP; second, the current study only used six WPs and the WP items used in the current study only matched two of the items used by Palm; and third, the current study used a within-subject comparison whereas Palm used a between-subject comparison.

Experiment 3 included three multiple-choice items, all of which could be correctly answered from the content in the word problem. The multiple-choice items were included as a way to ensure students were reading the problem text in full. Students performed well on these items and therefore it can be assumed that students were reading the problem text. However, the data from Experiment 3 did not support the findings reported by Palm (2008). It is possible that whatever effect the authentic cues may have had were diminished because of distractions incurred from the multiple-choice items. Students may have taken note of the authentic cues present in the problem text, but after reviewing three multiple-choice items and providing answered to those items, the information that was supposed to cue students had deteriorated and no longer had the intended effect. Students may have felt it unnecessary to read the problem text again when asked to solve the word problem.

The next major difference between Experiment 3 and Palm's (2008) work were the items chosen. The lack of effect in Experiment 3 may have been due to the items included. Not only did Experiment 3 use fewer WPs (e.g., six in comparison to 10), only two of the 10 items used by Palm (2008) – the Planks and Balloon WPs were included in Experiment 3. Regardless, our sample produced fewer RRs for the particular items that Palm used in his study.

Another possible explanation for the difference in RRs between Experiment 3 and Palm's (2008) work is that the current study mixed the standard and realistic WPs, while Palm used all realistic WPs. It could be that students in the current sample had a hard time switching back and forth between the realistic and standard WPs. Cognitive biases such as mental set (Luchins, 1942) may explain why students in Experiment 3 reported fewer RRs than the students in Palm's study. Students' ability to reason realistically may be dependent on the problems maintaining a similar format, otherwise students may use a strategy that is appropriate for one type of problem and apply that strategy to all problems. It is possible that improvement in RRs may depend on the WP type remaining consistent, and this is worth future examination.

The most consistent finding in the literature investigating student's ability to incorporate their real-world knowledge into the solution process is that of the increase of RRs to DWR WPs (De Corte et al., 1995; DeFranco & Curcio, 1993; Verschaffel et al., 2000). In all three experiments reported here the highest rate of RRs was for the DWR WP(s). However, the only instance when increased RRs to DWR WPs was found was an interaction effect in Experiment 1 with boys in the written response group providing more RRs to DWR WPs than boys in the control group. Boys have been found to benefit from stereotypical contexts (Boaler, 1994; Zohar & Gershikov, 2007) and the response sentence could have been an additional prompt to the context of the soldiers and busses

The results on gender differences in this line of research are sparse and mixed. Unfortunately, the results of the current study do not help to clarify the situation. The results of the current studies conflict, to some extent, with some of the literature on gender differences; in particular, girls tend to have a stronger need to understand math than boys (i.e., the strategy must appear to be sensible) (Boaler, 1994), girls are more likely to take context into account (Murphy & Whitelegg, 2006b, as cited in Palm & Nyström, 2009; Zohar & Gershikov, 2007), and girls experience more difficulty than boys in abstracting mathematical issues from their context (Murphy, 1990 as cited in Boaler, 1994). Arguably, the boys in our sample showed a stronger need to understand the math and that is reflected in the higher percentage of RRs overall, and that boys reported a higher percentage of RRs than girls on most of the realistic-items. It could also be that the experimental manipulation of using a response sentence may have been more salient for the boys than it was for the girls. The additional prompt could have caused the boys to evaluate the appropriateness of their numerical response in terms of the context of the problem.

There were some gender differences worth mentioning at the item level (see Appendix G). Boys provided more RRs than girls in response to the Rope WP (Experiment 1), the Temperature WP, and the Bus WP (Experiment 2, see Appendix C for WPs). While there have been reported findings of gender differences at the item level (see Boaler, 1993; Cooper & Harries, 2002, 2003; Palm & Nyström, 2009), these items were not the ones on which differences were expected. For example, Boaler (1993, 1994) reported that boys tend to be more engaged in the wood-cutting problem (the planks problem) at a deeper level than girls. That is, boys are more likely to create a situation model (e.g., drawing diagrams), whereas girls did not use a diagram at all and relied only on the numbers. There was no difference in the number of RRs to the planks item between boys and girls. The gender difference Cooper and Harries (2002, 2003) reported was in reference to a DWR WP (e.g., an elevator or lift), and while they reported no real gender difference, they did find that the boys were more likely than girls to “break the rules”, that is, they were more likely to surpass the restriction set to the number of people allowed on the elevator at one time. Palm and Nyström (2009) found gender differences pertaining to the running problem, a novel problem, and a horseback-riding problem. They found that boys performed better on the running problem, while girls performed better on the horseback-riding problem. Also, Palm and Nyström used a novel WP which has not been used in this line of literature. The horse-riding problem asks, “Elin is planning to ride horses each day for 4 days. Each day she has 45 minutes of free time to do this. How many 10-minutes rides does she have the time to do during these days?”, it was suggested that the context of this particular WP was stereotypical to girls and that explains why they performed better on that item (Palm & Nyström, 2009, p. 66).

Because gender differences resulted on items that have not previously shown gender differences speculations are made as to why these differences occurred. The difference on the Rope and Bus WP could be due to context (Boaler, 1993, 1994; Zohar & Gershikov, 2008). Perhaps boys have more experience tying ropes and their experiences beneficially contributed to providing a RR, and the Bus WP is about soldiers, which are stereotypically thought of as males. It may have been the stereotyped context that helped the boys and not the examples. As for the Temperature problem, the differences occurred in Experiment 2 when students were given examples, so it may be that the examples were of a greater impact among the boys and that is why they performed better than girls on those items.

Girls are more likely to take context into account, according to Murphy and Whitelegg (2006b as cited in Palm & Nyström, 2009). At the same time, Boaler (1993,

1994) states that girls are most likely to experience underachievement in math when they consider the implications of a context or situation. Together, the implication is that the kinds of WPs that require a student to consider the context are, by their very nature, going to facilitate poor performance among girls.

On the other hand, the girls who performed poorly in Boaler's (1993, 1994) study were responding to a standard word problem dressed up to impersonate a realistic word problem. The girls performed poorly on a fashion workshop word problem by including their real-world knowledge when they were not supposed to. The fashion workshop word problem was used as an analogous item for another abstract math problem that asked students to create combinations of numbers that equalled or were close to 24 from a list of numbers, without going over. The fashion workshop problem was similar to the abstract problem. The actions needed to perform each job (e.g., cutting fabric, sewing pieces of material together, and shipping) took a certain amount of time to conduct and there was only so much time in the work day. It was up to students to create combinations to get the most work done in the allotted time. Boaler (1993, 1994) found that the girls performed poorly on this problem in particular and suggested that it was because girls were considering issues such as, one has to cut fabric before it can be sewn and articles of clothing have to be complete before being shipped. Incorporating this knowledge into the solution process was disadvantageous to the girls.

The context of the fashion workshop word problem was flawed from a realistic perspective. The word problem mimics the types of realistic WPs used in the literature, but because the solution process ignored the realities of the possible combinations. In some ways, this problem is the opposite of a realistic WP. Because girls are found to be more likely than boys to take context into account (Murphy & Whitelegg, 2006b, as cited in Palm & Nyström, 2009) and have difficulty abstracting issues from the context (Boaler, 1994), they fell victim to erroneously employing their real-world knowledge when it was unwarranted. But taking context into account is what realistic problem solving is about. If this problem were to be called a realistic problem, and maybe it should be, then girls would be reported to be doing better on this problem than boys, rather than the reverse.

Girls did not, however, have difficulty abstracting issues from the wood-cutting or penalty-kicks word problems. This suggests that girls may be negatively impacted only by the realistic context if they see the relevance, that is, if the word problem has a stereotypical gendered context (see Cooper & Harries, 2002, 2003; Zohar & Gershikov, 2008). From this perspective, girls should be better able to incorporate their real-world knowledge on realistic word problems. However, the data reported in Experiments 1 to 3 do not support this supposition. More work needs to be conducted on gender difference in realistic word problems before any conclusive statements can be made.

Implications

The realistic word problems used in the current study reported here and within this line of literature are not the types of word problems that are being used in the educational system. Given students' lack of success on these problems, that is probably a good thing! Word problems are used in math classrooms to practice the basic arithmetic operations and to apply mathematical modelling to real-world situations (Greer, Verschaffel, & De Corte, 2003; Verschaffel et al. 2000). While standard word problems provide students the opportunity to practice basic arithmetic operations, they do not require students to evaluate their response as making sense in the context of the problem text. One could argue that realistic problems make use of applying mathematical reasoning to real-world situations tend to be more complicated than the application of one straightforward arithmetic operation. The low success rates on realistic WPs from this thesis, and the difficulty in improving them, reinforces the notion that word problem training by students today does not help them learn how to apply mathematical reasoning to real-world situations. In addition to identifying what does not work (Chapter 3), however, the results also support the importance of realistic reasoning (Chapter 2).

Practical Application. Standard word problems have already been shown to be a challenging task for students to complete correctly and that they perform quite poorly on word problems that require real-world consideration. However, our results indicate that the ability to correctly use one's real-world knowledge in realistic word problems is also an important skill in contributing to successful performance on standard word problems. This suggests that these types of problems should at the very least be considered when discussing how mathematical problem solving is relevant to the real-world.

It seems that the thread connecting standard and realistic word problems is that of creating a situation model. It may be that students who are providing more RRs to realistic WPs may be making use of building a situation model. Situation models are mental representations based on what the text is describing (Zwann & Radvansky, 1998). Students who are better able to create situations models perform better in standard WPs because those students are better able to translate the information into a mathematical model (Martin & Bassok, 2005). The students who are better able to solve realistic problems may be better able to mentally represent the problem situation which would seem to be a beneficial skill for problem solving in general.

When solving basic word problems, educators can also discuss how the numerical response could be impacted if consideration were given to the real-world implications of the situation. For example, the running question can be very easily and effectively displayed in real-time. Students can run a lap in a gymnasium while being timed using a stopwatch, then propose the question: how long it will take to run two laps, and so on. Questions can be posed as to why, at further distances, running time does not equate to the same proportional change in distance. One can also ask how the same question under a difference context can use the same proportional change rule (e.g., the boat question, see Appendix C). Posing these questions does not diminish the importance of the mathematical operation that is being practiced in the problem, instead it is bolstered by the computational rules' usefulness in predicting situations in our real-world under different situations.

Classroom Culture. It is important to recognize that the studies conducted in this dissertation and most in the literature place a large emphasis on word problem design as a way to improve realistic responding to realistic WPs (Dewolf et al., 2014, 2016; Palm, 2008; Xin et al., 2007). The implication is that factors such as word crafting and illustrations rather than teaching may be the key to WP performance. When discussing any skill taught in school, however, one cannot ignore the role played by teachers. Brousseau (1986) recognized that teachers and students are both members of an unspoken agreement, which he referred to as the didactic contract. Teachers and students engage in a set of reciprocal expectations and obligations which dictate how to behave in a math class, what kinds of problems to expect, and how to solve them.

Few studies have examined the instructional practices responsible for the development of reciprocal expectations. Verschaffel, De Corte, and Borghart (1997) found that preservice teachers (in their first and third year of study), similar to students, are resistant to include their real-world knowledge into the solution process. Teachers reported higher rates of RRs than students, but only report RRs half of the time (48%). A promising finding is that teachers in their third year of preservice education report more RRs than teachers in their first year.

When evaluating student responses to realistic WPs, preservice teachers rated non-realistic answers as correct in comparison to RRs in 80% of cases (Verschaffel et al., 1997). Again, teachers in their third year of training gave more partial marks to students for RRs to realistic WPs. The authors suggest that teachers have a strong tendency to model and interpret school arithmetic WPs in a non-realistic way and these beliefs have an impact on their actual teaching practices and subsequently on their students' learning processes and outcomes.

These differences in preservice teachers' evaluation of student responses to realistic WPs between first and third year of training (Verschaffel et al., 1997) suggests that, with time, teachers may 'grow out' of their resistance towards real-world knowledge being used in the solution process. However, when teachers solve non-standard word problems jointly with their students, teachers continue to solve problems in a *superficial* way (Rosales, Vicente, Chamoso, Muñez, & Orrantia, 2012). Rosales and colleagues describe the superficial problem-solving strategy as not building a situation model of the problem and the mathematical model is not based on mathematical reasoning. The authors note that in some cases, when the situation model did not fit the mathematical model, teachers recreated or ‘twisted’ the situation model to fit the mathematical model being used.

The literature on student and teacher performance on realistic WPs paints a picture of cyclical experiences. Students experience the standard formulae for solving WPs during their educational careers and some students continue to receive post-secondary training as teachers. Preservice education to preservice teachers continues to promote superficial strategies for solving WPs (Rosales et al., 2012; Türker, Sağlam, & Umay, 2010; Verschaffel et al., 1997) and this style is disseminated when teachers are then in a position to teach students. As results from this thesis and from other research strongly suggest that WP design is not improving students’ ability to reason realistically, preservice education programs and professional development opportunities that support teachers in incorporating the consideration of realistic factors in problem solving may be the best way to advance mathematical pedagogy for word problems solving. Such a pedagogical approach could change the expectations around word problems, and specifically change the didactic contract.

Limitations

Some of the biggest issues faced in these series of experiments were to do with the randomization process and measurement errors. The first two studies were designed with randomization at the school level rather than at the individual level, and the small number of schools meant that our randomization process did not result in comparable groups. For anyone who conducts research in the public-school system, the primary challenge is to balance the time necessary to complete the study with the time taken away from classroom activities. As such, it was decided that it made more sense to randomly assign schools to conditions, rather than assign individuals to conditions. This strategy was especially relevant in the first experiment where there were many trips required for testing and a large amount of time for individual testing using the WJ III Ach (e.g., up to 90-minutes). Because groups were defined by the set of instructions they received, and instructions were given at the group level, group administration ended up effectively meaning that randomization was at the school level. In retrospect, it may have been better to randomly separate each school into two groups that would then be assigned to each condition. Although testing two groups in each school would have required greater schedule flexibility and more rooms than were available at the participating schools, it may also have mitigated some of the initial group differences. As it was, there were significant individual differences between conditions that randomizing by individual should have eliminated. Similar individual differences were found between conditions in the second experiment on the Raven's scores. When random assignment of individuals to conditions was possible, as was done in Experiment 3, the randomization process did its job and there were no individual differences between conditions prior to the experimental manipulation.

While there are obvious reasons for counter-balancing, in hindsight, it may have been better to choose an order that would have facilitated learning instead of mixed counter-balanced ones. Instead, items could have been provided according to difficulty level with easier WPs placed at the beginning of a test booklet and harder WPs placed near the end of a test booklet. Having the booklets set up that way may have allowed students' confidence to be built early in the testing process. Placing the easier items near the beginning could provide students with a sense of self-competence that can facilitate the motivation needed when faced with more challenging problems. You can imagine the dismay of a student who gets challenging items first and then feels as though they may not have the skills necessary to complete the booklet, when in fact, there are easier items further along. This can result in the student simply shutting down and submitting an incomplete test booklet.

Future Directions

Given the results presented above, there are many different avenues for future research. The possibilities suggested here can be broadly grouped into two categories that parallel the approach taken in this thesis: a) exploring further ways to investigate individual differences; and b) exploring other ways to improve realistic responding.

Not all students perform poorly on realistic WPs, and more research needs to be conducted on which factors (e.g., reading ability, gender, math ability) contribute to realistic reasoning in math. Within the research on standard word problems, there is evidence of individual differences in performance. Evidence from the work of Hegarty, Mayer, and Monk (1995) shows that what good problem solvers and poor problem solvers look at in a word problem varies. Hegarty and colleagues (1995) monitored students' eye fixations as they responded to mathematical word problems, and they found that good problem solvers spent more time looking at variable names than poor problem solvers. By focusing on how pieces of information were connected, good problem solvers construct a meaningful representation of the situation described. On the other hand, poor problem solvers spent more time looking at numbers and relational terms (e.g., keywords like, "*less*" or "*more*"). The approach taken by the poor problem solvers was considered a direct translational approach (Hegarty et al., 1995) where students take the numbers and apply a mathematical operation that is deemed appropriate based on the keywords. Perhaps those who perform well on realistic items are paying more attention to information in the problem that helps them build a mental representation, or a situation model. An eye-tracking study with realistic word problems might demonstrate similar differences between those who are good at realistic word problems and those who are not, or those could demonstrate how good realistic problem-solvers might be different than good standard problem-solvers.

It could be that students who incorporate their realistic knowledge into the classroom when it is unnecessary are being penalized for doing so (see Verscahffel et al. 1997). Although this would not be true in all teachers' classrooms, there may be students who receive poorer grades for not providing the 'correct' numerical answer, and they may also be singled out as trouble-makers who are always trying to avoid 'answering' the problem. In this scenario, students who use their realistic knowledge in the classroom are doing so at the detriment of their formal performance (e.g., grades).

On the other hand, those students who use their real-world knowledge to help solve these problems may not be interested in conforming to the didactic contract. If this were the case, the students who understand that their real-world knowledge is important and necessary for mathematical modelling may be the same students who get discouraged, perform poorly, discontinue their math education at the high-school level, drop out of school, or never go onto post-secondary school. Hopefully, this is not the case, and the formal education system is not deterring individuals from using their applied sense of how mathematical modelling is used (e.g., considering how math works in the real-world).

Furthermore, given the data in this dissertation that suggests that the ability to provide realistic responses predicts success on standard WPs, there is yet another reason to try and encourage the development of mathematical modelling of real-world situations. Nevertheless, it would be worthwhile for future research in individual differences to see if those who use realistic knowledge are those who are especially smart, non-conformist, or show some other signs of academic rebellion that might negatively affect their attitude towards school. There may even be different sub-types of children who use realistic knowledge.

One of the issues with the realistic WPs typically used in the literature is that the problems cover a variety of mathematical operations. This matters because it may be the case that different kinds of problems will require different kinds of modifications. In the research on standard WPs, researchers have identified three distinct classifications for word problems involving addition and subtraction: (1) Change; (2) Compare; and (3) Combine type problems (Verschaffel & De Corte, 1997a). Multiplication and division problems are not as nicely broken up, but include classification types such as Equal Groups, Equal Measures, Rate, Measurement Conversion, Cartesian Product, etc. (Verschaffel & De Corte, 1997a). Not only are there different classification systems for problem type, the difficulty level also varies. With multiplication and division problems, Equal Groups and Measures are considered to be somewhat easy as students perform well on these problems, while Cartesian Product and Measurement Conversion are considered rather difficult as students do not perform as well on these problems (Verschaffel & De

Corte, 1997a). Within addition and subtraction problems, the relation is more complex. While there are three types of problems (e.g., Change, Compare, and Combine), there are six subtypes within each of the three broad types of addition and subtraction problems. Each subtype involves three different unknowns, the Result Set (the sum or difference), the Change Set (how much was added or subtracted from the original or start quantity), or the Start Set (what was the original or starting quantity) and each involves an increase or a decrease. Moreover, the difficulty level also varies between the six subtypes. For example, within the Change problem type, students experience the most difficulty when the initial quantity is unknown (e.g., Start Set problems) compared to problems when then final sum or difference is unknown (e.g., Result Set problems) (Verschaffel & De Corte, 1997a). The complexities of standard WPs may appear to be unrelated to applying realistic knowledge to them; however, it raises the possibility that similar challenges may be faced with realistic problems, and different pedagogical changes may be required for each.

Future studies could aim to increase realistic responses at the item level, rather than one intervention aimed at increasing realistic responses for all realistic WPs in general. The response sentence had a positive effect on the number of realistic responses provided for DWR WPs, but only for boys. When students report a fractional numerical response within the context of the problem (e.g., half a bus, or half a balloon), it is reasonable to see why a response sentence may work to prompt students as to whether their numerical answer makes sense given the situation described in the word problem. However, using this response sentence on WPs like the running item does not provide a prompt that something about the numerical answer does not fit with the situation described in the problem. Similarly, the response sentence would not provide a prompt for other types of problems that ask the number of guests at the birthday party, or how old Rob is.

In many cases, new pedagogical approaches to teaching students to enhance their ability to solve realistic word problems. at the item level can be inexpensive and easy to employ. For instance, consider the planks problems, “Steve has bought 4 planks of 2.5 m each. How many planks of 1m can he get out of these planks?” (Verschaffel et al., 1994, p. 276). This problem could be easily done as a hands-on activity. Students can be given wooden skewers with the sharp end cut off, where 5cm is equivalent to 1m, and they are asked to make use of the skewers as a model for the problem. Students can use scissors to cut the skewers at the 1m mark and they will be able to see that the pieces left over are not able to make a 1m plank. Now, one could imagine students coming up with creative ideas to put together those pieces, such as, using glue or tape to put them together.

The experiments conducted in the current dissertation (Chapter 3) were not designed around a theoretical framework and instead were designed around questions that stemmed from the work of other researchers. Under the theoretical framework of the didactic contract, future work would do well to include teachers, as teachers are the other participants in the didactic contract. A consistent argument presented in this line of literature is that students are resistant to including their real-world knowledge into the solution process because they know from their experience in the classroom that using that information is not expected of them (i.e., from students’ role in the didactic contract). Conversely, teachers do not expect their students to incorporate realistic reasoning when solving math word problems and teachers are expected to grade the numerically correct response given to any problem (i.e., teachers' role in the didactic contract). Therefore, if both teachers and students are performing under the notion of these preconceived roles, and it is hypothesized that these pre-existing roles make it difficult to induce realistic responding in students, the next obvious step is to change how they are expected to behave in their respective roles by altering the didactic contract.

It seems likely that increasing RRs among elementary school students would be more successful if teachers were included in the process. Future studies should be aimed at changing teachers' attitudes towards using realistic reasoning in the math classroom. The first step would be an intervention study that involved the teachers demonstrating to students how to solve realistic word problems. Future studies could then examine different pedagogical methods used by teachers to incorporate realistic reasoning in the math classroom. This could be achieved through a series of experiments training teachers on how to incorporate realistic reasoning into their classrooms in general. Alternatively, this information could also be disseminated via professional development days, where researchers provide teachers with a workshop on realistic reasoning. Another possibility would be to target teachers during their preservice education programs, thereby adjusting new educators' roles within the didactic contract before becoming immersed in the classroom culture.

Another way to lessen the 'suspension of sense making' too commonly found in elementary school students would be to talk more about math and its real-world application. When discussing a straightforward word problem, mathematical reasoning can be promoted by posing questions that challenge student's traditional thinking. For example, one could ask students to brainstorm situations that would result in the numerical answer as being incorrect, or imagine situations that would result in the numerical answer being different than what the computation would suggest. Students could be encouraged to consider if the problem posed in the text were something encountered in our everyday lives, and if the problem solution would be the same or different in that context.

Having students work in groups to discuss both standard and realistic word problems may be one type of effective pedagogical approach. In a typical mathematical classroom in North America, teachers present students with the procedures and examples to solve a problem, and then students work individually on a series of problems from the textbook. The content that is not finished in class can be given as homework, and the classroom discussion can be about going over answers and mistakes (Sheie, 2016). Instead, students should be more active in their learning. There should be opportunities to ask many questions and consider situations when procedures learned may not be appropriate, and how the problem could be approached. Students need to be able to talk about their findings, especially if they feel that their findings may not make sense. This can be empirically tested by randomly assigning a classroom to have lessons plans about word problems designed around group discussion of the problem and its situation and compare that to a business-as-usual class. There are opportunities everywhere for illustrating mathematically modelling, but teachers need to be able to recognize them and feel comfortable approaching them (another research area completely)!

Conclusion

Students often ask when they are ever going to use their math skills after they become adults and are living beyond the world of primary and secondary educational institutions. The answer is every day! Students must recognize that they will be making mathematical based judgements daily. Mathematics is so amazing partly because it is so useful in our everyday lives. The focus of this thesis on realistic problem solving is that students see formal mathematics as quite the opposite – useless. Although the data presented here suggests that teaching students how to think about mathematical problem solving in a realistic way is difficult to do, being able to demonstrate the usefulness of mathematics to our everyday lives makes it well worth the effort to try to do so.

References

Bjork, I. A., & Bowyer-Crane, C. (2013). Cognitive skills used to solve mathematical word problems and numerical operations: A study of 6- to 7-year-old children. *European Journal of Psychological Education, 28*, 1345-1360. doi:10.1007/s10212-012-0169-7

Boaler, J. (1993). Encouraging the transfer of 'school' mathematics to the 'real world' through the integration of process and context, context and culture. *Educational Studies in Mathematics, 25*, 341-373.

Boaler, J. (1994). When do girls prefer football to fashion? An analysis of female underachievement in relation to 'realistic' mathematic contexts. *British Educational Research Journal, 20*, 551-564.

Brousseau, G. (1986). The didactic relationship: The middle, *Proceedings of the Fourth Summer School of Didactics Mathematics, pp. 54-68*, IREM Paris 7.

Cooper, B., & Dunne, M. (2000). *Assessing children's mathematical knowledge: Social class, sex and problem-solving*. Buckingham, Philadelphia: Open University Press.

Cooper, B., & Harries, T. (2002). Children's responses to contrasting 'realistic' mathematics problems: Just how realistic are children ready to be? *Educational Studies in Mathematics, 49*, 1-23.

Cooper, B., & Harries, T. (2003). Children's use of realistic considerations in problem solving: Some English evidence. *Journal of Mathematical Behavior, 22*, 451-465. doi:10.1016/j.jmathb.2003.09.004

De Corte, E., Verschaffel, L., & Lasure, S. (1995, April). *Word problems: Game or reality? Studies of children's beliefs about the role of real-world knowledge in mathematical modeling.* Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, US.

DeFranco, T. C., & Curcio, F. R. (1993). A division problem with a remainder embedded across two contexts: Children's solutions in restrictive vs. real-world settings. *Focus on Learning Problems in Mathematics, 19*, 58-72.

Dewolf, T., Van Dooren, W., Ev Cimen, E., & Verschaffel, L. (2014). The impact of illustrations and warnings on solving mathematical word problems realistically. *The Journal of Experimental Education 82*, 103-120. doi:10.1080/00220973.2012.745468

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., …Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29-43. doi:10.1037/002-0663.98.1.29

Greer, B. (1993). The mathematical modeling perspective on wor(l)d problems. *Journal of Mathematical Behavior, 12*, 239-250.

Greer, B., Verschaffel, L., & De Corte, E. (2003). "The answer is really 4.5": Beliefs about word problems. In G. Leder, E. Pehkonen, & G. Torner (Eds.), *Beliefs: A hidden variable in mathematics education* (pp. 271-292). Secaucus, NJ, US: Kluwer Academic Publishers. Retrieved from http://www.ebrary.com

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.

Inoue, N. (2008). Minimalism as a guiding principle: Linking mathematical learning to everyday knowledge. *Mathematical Thinking and Learning, 10*, 36-67. doi:10.1080/10986060701820293

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

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs, 54*(6), i-95. http://dx.doi.org/10.1037/h0093502

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

Newfoundland Labrador, Education and Early Childhood Development. (2015). *Mathematics 6 Curriculum Guide 2015.* Retrived from http://www.ed.gov.nl.ca/edu/k12/curriculum/guides/mathematics/grade6/Mathematics_Grade6_Curriculum_Guide_final.pdf

Pimperton, H., & Nation, K. (2010). Understanding words, understanding numbers: An exploration of the mathematical profiles of poor comprehenders. *British Journal of Educational Psychology, 80*, 255-268. doi:10.1348/000709909X477251

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 7*, 37-58. doi:10.1007/s10649-007-9083-3

Palm, T., & Nyström, P. (2009). Gender aspects of sense making in word problem solving. *Journal of Mathematical Modelling and Application, 1*, 59-76.

Reusser, K. (1988). Problem solving beyond the logic of things: Contextual effect on understanding and solving word problems. *Instructional Science, 17*, 309-338.

Reusser, K., & Stebler, R. (1997). Every word problem has a solution – The social rationality of mathematical modeling in schools. *Learning and Instruction, 7*, 309-327.

Silver, E. A., Shapiro, L. J., & Deutsch, A. (1993). Sense making and the solution of division problems involving remainders: An Examination of middle school students' solution processes and their interpretations of solutions. *Journal for Research in Mathematics Education, 24*, 117-135.

Sheie, A. H. (2016). *Research based best practices for teaching mathematics and improving math attitudes* (Honors Senior Theses). Retrieved from http://digitalcommons.wou.edu/honors_theses

Schumacher, R. F., & Fuchs, L. S. (2012). Does understanding relational terminology mediate effects of intervention on compare word problems? *Journal of Experimental Child Psychology, 111*, 607-628. doi:10.1016/j.jecp.2011.12.001

Verschaffel, L., & De Corte, E. (1997a). Word problems: A vehicle for promoting authentic mathematical understanding and problems solving in the primary school? In *Learning and Teaching Mathematics.* T. Nunes, & P. Bryant (Eds.) (pp. 69-97). East Sussex, UK: Psychology Press Ltd.

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4*, 273-294.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making Sense of Word Problems.* Lisse, The Netherlands: Swets & Zeitlinger.

Vilenius-Tuohimaa, P., Aunola, K., & Nurmi, J-E. (2008). The association between mathematical word problems and reading comprehension. *Educational Journal, 28,* 409-426. doi:10.1080/01443410701708228

Walkington, C., Sherman M., & Petrosino, A. (2012). “Playing the game” of story problems: Coordinating situation-based reasoning with algebraic representation. *Journal of Mathematical Behavior, 31*, 174-195. doi:10.1016/j.jmathb.2011.12.09

Weyns, A., Van Dooren, W., Dewolf, T., & Verschaffel, L. (2016). The effect of emphasizing the realistic modelling complexity in the text or picture on pupils’ realistic solutions of P-items. *Educational Psychology, 37*, 1173-1185. doi:10.1080/01443410.2016.1259461

Xin, A., Lin, C., Zhang, L., & Yan, R. (2007). The performance of Chinese primary school students on realistic arithmetic word problems. *Educational Psychology in Practice: Theory, Research and Practice in Educational Psychology, 23* 145-159. doi:10.1080/02667360701320853

Xin, A., & Zhang, L. (2009). Cognitive holding power, fluid intelligence, and mathematical achievement as predictors of children’s realistic problem solving. *Learning and Individual Differences, 19*, 124-129. doi:10.1016/j.lindif.2008.05.006

Yoshida, H., Verschaffel, L., & De Corte, E. (1997). Realistic considerations in solving problematic word problems: Do Japanese and Belgian children have the same difficulties? *Learning and Instruction, 7*, 329-338.

Zohar, A., & Gershikov, A. (2008). Gender and performance in mathematical tasks: Does the context make a difference? *International Journal of Science and Mathematics Education, 6*, 677-693.

Zwann, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.

Appendices

Appendix A

Informed Consent

Supervisor: Darcy Hallett, Associate Professor of Psychology
Email: darcy@mun.ca
Phone: 709-864-4871, Fax: 709-864-2340

Cheryll Fitzpatrick, Ph.D. Candidate
Email: cheryllf@mun.ca, Phone: 709-864-3287

Greetings,

Your child's school has been invited to take part in a research project investigating children's knowledge and understanding of mathematical word problems.

This form is part of the process of informed consent. It will provide you with information regarding what the research is about and what your child's participation will involve. It also describes your child's right to withdraw from the study at any time. In order to decide whether you wish to have your child participate in this research study, you should understand enough about its risks and benefits to be able to make an informed decision. Take time to read this carefully and to understand the information given to you. Please contact the researcher, Cheryll Fitzpatrick if you have any questions about the study.

It is you and your child's decision whether to take part in this research. If you or your child chooses not to take part in this research or if you or your child decides to withdraw from the research once it has started, there will be no negative consequences for your child, now or in the future.

Introduction:

Mathematical word problems, or story problems, have become the most common kind of problem found in formal education. Yet, the National Mathematics Advisory Panel (U.S. Department of Education, 2008) reports that word problems are one of the three areas for which students have the poorest preparation. My goal is to find ways that make solving word problems easier for children.

Purpose of the study:

This study's main purpose is to determine if students perform better on mathematical word problems when they have been given an opportunity to answer non-mathematical questions about the word problem before having to perform any of the math needed to solve the problem.

What will your child do in this study?

If your child participates in this study, he or she will complete a few tasks. Your child will participate in various tasks that will take place in a group, with the rest of their classmates, and in one-on-one interviews with the researcher. Group activities will consist of a pattern completion task and solving mathematical word problems. Individual activities will consist of reading comprehension and mathematical abilities tasks and at the very end an interview to discuss their responses on the mathematical word problems that were already solved.

Length of time:

This study will occur over a series of sessions (e.g., over a few days), in total the entire study will take your child 3.5-4 hours to complete.

Withdrawal from the study:

Your child is free to withdraw from the study with no consequences. If your child decides to no longer be involved in the study, he or she can inform the researcher during or after the interview. In the case of withdrawal, the related data will be disposed of prior to data analysis.

Possible benefits:

This will give your child extra practice with math word problems. In addition, your child will have an opportunity to experience scientific research first-hand and contribute to the advancement of the field.

Possible risks:

The only conceivable risks are test and math anxiety, even though this is not a test and will not count towards school grades.

Confidentiality:

Results for each child are kept strictly confidential. Every reasonable effort will be made to assure your child's anonymity. If the results are published in a scientific journal, it will be two or three years after the end of the study. Summaries of information about different groups of participants will be given. There will be no permanent record kept that your child participated in this study.

Your child's data will be stored at Memorial University of Newfoundland's Research Centre for the Development of Mathematical Cognition. Data will be stored in a secured area in this locked laboratory in which only those associated with the study have access. Electronic data will also be stored on a computer that is password protected in the locked facility. Data will be stored for a minimum of five years, as required by Memorial University policy on Integrity in Scholarly Research.

Supervisor: Darcy Hallett, Associate Professor of Psychology
Email: darcy@mun.ca
Phone: 709-864-4871, Fax: 709-864-2340

Cheryll Fitzpatrick, Ph.D. Candidate
Email: cheryllf@mun.ca, Phone: 709-864-3287

- Your child's school has agreed to take part in a study through Memorial University that is designed to investigate children's knowledge and understanding of mathematical word problems
- Children will be visited by researchers in their school, and will be asked to complete a series of tasks
- Participation will have no effect on your child's school grades
- If you have any additional questions that are not answered by the information sheet, please contact Dr. Darcy Hallett, or Cheryll Fitzpatrick through the contact information listed above
- The proposal for this research has been reviewed by the Interdisciplinary Committee on Ethics in Human Research and found to be in compliance with Memorial University's ethics policy. If you have ethical concerns about the research, such as the way you have been treated or your rights as a participant, you may contact the Chairperson of the ICEHR at icehr@mun.ca or by telephone at 709-864-2861
- Please fill out the form below to indicate whether or not you would like your child to participate

Your signature on this form means that:

- ❖ You have read the information about the research.
- ❖ You understand what the study is about and what your level and your child's level of involvement are.
- ❖ You understand that any data collected from you or your child up to the point of your withdrawal will be destroyed.
- ❖ You understand that you are consenting to the use of the data provided by yourself on the demographics sheet accompanying this consent form

Your signature:

I have read and understood what this study is about and appreciate the risks and benefits. I have had adequate time to think about this and I agree to participate voluntarily and I understand that I may end my or my child's participation at any time.

____
Your child's name

______________________________ ______________________________
Signature of Guardian Date

Appendix B

Demographics Form

Questions for children to answer.

What is your full name?

When is your birthday?

Are you a boy or a girl, circle one: Boy Girl

Do you have any brothers or sisters? If yes, how many?

Who is your best friend or best friends?

What do you like to do after school?

Do you play any sports or musical instruments? If so, what do you play?

What is your favourite school subject?

Is there anything else interesting about yourself that you would like to tell me?

Appendix C

Standard WPs	Realistic WPs
Pete organizes a birthday party for his tenth anniversary. He invited 8 boy friends and 4 girl friends. How many friends did Pete invite for his birthday party?	Carl has 5 friends and Georges has 6 friends. Carl and Georges decide to give a party together. They invite all their friends. All friends are present. How many friends are there at the party?
A shop-keeper has two containers for apples. The first container contains 60 apples and the other 90 apples. He puts all apples into a new, bigger container. How many apples are there in the new container?	What will be the temperature of water in a container if you pour 1 L of water at 80° and 1 L of water of 40° into it?
Steve has bought 5 planks of 2 m each. How many planks of 1 m can he saw out of these planks?	Steve has bought 4 planks of 2.5 m each. How many planks of 1 m can he get out of these planks?
Pete's piggy bank contains 690 franks. He spends all that money to buy 20 toy cars. How much was the price of one toy car?	450 soldiers must be bussed to their training site. Each army bus can hold 36 soldiers. How many buses are needed?
A boat sails at a speed of 45 km/hr. How long does it take this boat to sail 180 km?	John's best time to run 100m is 17 sec. How long will it take to run 1 km?
Chris made a walking tour. In the morning he walked 8 km and in the afternoon he walked 15 km. How many kilometers did Chris walk?	Bruce and Alice go to the same school. Bruce lives at a distance of 17 km from the school and Alice at 8 km. How far do Bruce and Alice live from each other?
Kathy, Ingrid, Hans and Tom got a box containing 14 chocolate bars from their grandfather, which they share equally among them. How many chocolate bars does each grandchild get?	Grandfather gives his 4 grandchildren a box containing 18 balloons, which they share equally. How many balloons does each grandchild get?
This morning Steve had 480 francs in his piggy bank. Now he has already has 1650 francs in his piggy bank. How many francs did Steve gain since this morning?	Rob was born in 1978. Now it's 1993. How old is he?
A man cuts a clothesline of 12 m into pieces of 1.5 m each. How many pieces does he get?	A man wants to have a rope long enough to stretch between two poles 12 m apart, but he has only pieces of rope 1.5 m long. How many of these pieces would he need to tie together to stretch between the poles?
A shop sells 312 birthday cards in December. About how many do you think it will sell altogether in January, February and March?	A shop sells 312 Christmas cards in December. About how many do you think it will sell altogether in January, February and March?

Appendix D

Pair 1

S1 Pete organizes a birthday party for his tenth anniversary. He invited 8 boy friends and 4 girl friends. How many friends did Pete invite for his birthday party?

Answer:

Comments:

P1 Carl has 5 friends and George has 6 friends. Carl and George decide to give a party together. They invite all their friends. All friends are present. How many friends are there at the party? (Nelissen, 1987, as cited in Verschaffel et al., 1994)

Answer:

Comments:

Appendix E

Post-WP interview questions (Palm, 2008):

1. Describe your thinking on which you based your solution.
2. Did you think of anything else, and did you make any other considerations, during the solution process?
3. A question about whether the students' had made relevant realistic considerations during the solution process, explicitly pointing to a task specific phenomenon. For example; did you think about whether the runner could keep the same speed during the whole race?
4. Why did you not pursue this line of thinking in your solution to the task? (if the student had made realistic considerations but not used them in the solution to the task)
5. A task specific question aimed at finding out whether the students' had the real-world knowledge required for solving the task successfully. Concerning the running task this question was: Do you think the runner can keep the same speed during the whole race? (p. 68)

Appendix F

The original planks problem from Verschaffel et al. (1994):

> Steve has bought 4 planks of 2.5 m each. How many planks of 1 m can he saw out of these planks? (p. 276)

The modified, enhanced authenticity, version of the planks problem from Palm (2008):

> [Steve is] building a cabin and as walls want to use planks that are 1 m long. [He] are at the moment short of thirteen 1-meter planks. A friend says that she has found 4 planks, each 2.5 m long. [Steve is] wondering if that is enough to finish the walls. How many 1-meter planks can [he] saw out of the planks she found? (p. 44)

1. Why is Steve using a saw?
   a) To put the 5 pieces of planks together
   b) To cut the planks into 1 m pieces
   c) To cut the planks into 2 m pieces
   d) To put the 2 pieces of planks together

2. What is a plank?
   a) A type of small organism that lives in the water
   b) A long single piece of wood
   c) An exercise position
   d) None of the above

3. Why is Steve sawing the planks?
   a) Steve is making a fence
   b) Steve is making a bookshelf
   c) Steve just likes to saw things
   d) Steve is a pirate and needs many planks

4. Answer the above math problem in a complete sentence.

Appendix G

Supplementary Statistical Analyses – Chapter 2

The WCJ III Ach is a well validated and standardized assessment. To determine if the current sample was performing based on age-norms (see WJ III Ach Normative Update Technical Manual, McGrew, Schrank, & Woodcock, 2007), their scores across each composite measure (e.g., Math Calculation, Reading Comprehension, and Math Reasoning) and subtest were evaluated (e.g., Math Calculation Skill, Math Fluency, Passage Comprehension, Vocabulary, Applied Problems, and Quantitative Concepts) using a series of single-sample t-tests. We found that the current sample of students are performing significantly below age-norms on the Reading Comprehension and Math Calculation composite measures [$t(71) = -3.364$, $p =.001$, Cohen's $d = .268$, and $t(71) = -7.383$, $p < .0005$, Cohen's $d = .691$, respectively], and no difference on scores for Math Reasoning, $t(71)$ 1.806, $p = .075$.

To break these analyses down further, the norms for the subtests of the Math Calculation and Reading Comprehension composite scores were evaluated. Student's below average scores on the Math Calculation composite measure is being driven by their subpar performance on the Math Calculation Skills and Math Fluency subtests, [$t(71) = -7.583$, $p < .0005$, Cohen's $d = .673$, and $t(71) = -4.973$, $p < .0005$, Cohen's $d = .467$, respectively]. Student's below average scores on the Reading Comprehension composite measure were being driven by poor performance on the Passage Comprehension subtest, $t(71) = -6.009$, $p < .0005$, Cohen's $d = .432$, and no difference was found on the Vocabulary subtest (Table 1).

Table 1

Mean Scores (SD) on WJ III Ach Subtests and Percentage of Students Below the Age Norm (N = 72)

	Subtest Measure					
	CS	MF	PC	Voc.	AP	QC
Mean (SD)	86.83 (11.25)	91.53 (12.09)	92.46 (7.97)	98.06 (11.04)	99.46 (9.41)	101.47 (13.38)
% Below Age Norm	86.1	73.6	86.1	56.9	44.4	47.2

Note. CS = Calculation Skill; MF = Math Fluency; PC = Passage Comprehension; Voc. = Vocabulary; AP = Applied Problems; QC = Quantitative Concepts.

In Chapter 2, a hierarchical regression analysis was used to demonstrate that individual differences (e.g., gender, Raven's Matrices score, and composite measures on the WJ III Ach) were predictors of performance on S-items. The WJ III Ach composite measure, Mathematical Reasoning, remained a unique predictor of performance on S-items after everything was entered into the model. To determine which of two components of Mathematical Reasoning (e.g., Quantitative Concepts and Applied Problems), or perhaps both, may have been driving the relation between math reasoning and standard WP solving performance, a regression analysis was conducted where performance on the S-items was entered as the criterion, and scores on Raven's Matrices, all six subtests of the WJ III Ach, Gender, and RRs were predictors. All predictors accounted for a significant amount of the variance in accuracy on S-items, $F(9, 62) = 9.773$, $p < .0005$ (Table 2) and it is the Quantitative Concepts subtest of Math Reasoning composite score (in addition to RRs) that continues to add a unique amount of variance.

Table 2

Predictors of Accuracy on Standard WPs

Variable	R^2	*B*	*SE (B)*	β
Constant		-6.122	1.592	
Raven's Matrices		.054	.034	.148
Calculation Skill		.006	.013	.053
Math Fluency		.000	.012	-.002
Passage Comprehension		.025	.016	.161
Vocabulary		-.013	.013	-.117
Applied Problems		.013	.016	.098
Quantitative Concepts		.045**	.014	.490
Gender		.198	.230	.080
RRs		.348**	.118	.259
Total Model	.587***			

Note. *B* denotes the variable estimate, *SE B* denotes the standard errors of the variable estimate, β denotes the standardized estimate, and * $p < .05$, ** $p < .01$, *** $p < .001$

<u>**Supplementary Statistical Analyses – Chapter 3**</u>

While examining differences in at the item level was not part of the primary analyses in the current studies, we have included this supplementary information because other researchers have indicated some items may elicit more RRs than others, specifically DWR WPs. In Experiment 1, using a 2 (Condition [Control, Written Response] x 2 (Gender) ANCOVA controlling for scores on Math Reasoning we found a main effect of Condition for the Planks WP, $F(1, 67) = 5.126$, $p = .027$, $\eta_p^2 = .071$, with unexpectedly higher RRs occurring in the control condition. There was also a main effect of Gender for the Rope WP, $F(1, 67) = 6.296$, $p = .015$, $\eta_p^2 = .086$, where boys were providing more RRs than girls. Lastly, there was a marginal Condition by Gender interaction, again for the Rope WP, $F(1, 67) = 3.961$, $p = .051$, $\eta_p^2 = .056$, with boys in the control group providing more RRs than boys in the written response group (see means in Chapter 3 Table 2).

In Experiment 2, using a 2 (Condition [Written Response, Examples] x 2 (Gender) ANCOVA controlling for scores on Raven's Matrices, we found a main effect of Condition for the School WP, $F(1, 134) = 7.300$, $p = .008$, $\eta_p^2 = .052$, with a higher rate of RRs in the written response condition. There was a main effect of Gender for the Temperature and Bus WPs [$F(1, 134) = 8.855$, $p = .003$, $\eta_p^2 = .062$ and $F(1, 134) = 9.847$, $p = .002$, $\eta_p^2 = .068$, respectively], with boys giving higher rates of RRs than girls, in both WPs. Lastly, there was a Condition by Gender interaction for the Temperature WP, $F(1, 134) = 6.089$, $p = .015$, $\eta_p^2 = .043$, with boys in the written response condition having a higher rate of RR than boys in the examples condition (see means in Chapter 3, Table 4).

In Experiment 3, using a 3 (Condition [Control, Written, Response, Examples]) x 2 (Gender) ANOVA we found no effect of Condition, Gender, or a Condition by Gender interaction on any of the individual WPs.