

# The Effect of Contextual and Conceptual Rewording on Mathematical Problem- Solving Performance

Majid Haghverdi and Lynda R. Wiest

*This study shows how separate and combined contextual and conceptual problem rewording can positively influence student performance in solving mathematical word problems. Participants included 80 seventh-grade Iranian students randomly assigned in groups of 20 to three experimental groups involving three types of rewording and a control group. All participants completed a pretest. The three experimental groups completed posttests involving problem rewording, and the control group received a posttest identical to the pretest. Rewording in general was shown to decrease student errors in solving word problems. There was no significant difference among contextual, conceptual, and combined contextual/conceptual rewording in reducing errors. In terms of Knifong and Holtan's (1976) classification of word problem errors, the rewordings had no significant effect on decreasing "clerical and computational" errors, but they significantly reduced "other errors" (such as use of wrong procedures or operations or providing no response).*

---

---

Word problems are an important part of the mathematics curriculum at all levels (Cai & Lester, 2010; Mahajne & Amit, 2009). Their use in mathematics instruction is intended to contextualize mathematics in order to help students make better sense of the embedded mathematics concepts, to apply mathematics concepts and procedures to real-world situations, and to motivate students (Depaepe, De Corte, & Verschaffel, 2010; Walkington, Sherman, & Petrosino, 2012). Despite their prominent role in school mathematics, students tend to fear and perform poorly on mathematical word problems (Ahmad,

---

---

**Majid Haghverdi** Majid Haghverdi is an assistant professor at Islamic Azad University, Arak Branch. His research interests include problem solving and teacher education.

**Lynda R. Wiest** is a professor at the University of Nevada, Reno. Her scholarly interests include mathematics education, educational equity, and teacher education.

Salim, & Zainuddin, 2008; Awofala, 2011). They have been shown to need support bridging word problem meaning to the formal and informal mathematical processes needed to solve word problems (Walkington et al., 2012), and it is uncertain whether students successfully transfer their mathematics knowledge from school to everyday settings (Depaepe et al., 2010).

Several aspects of word problems, such as their context and their linguistic structure, have been suggested as factors that influence problem-solving performance. Because studies of potential factors have yielded inconclusive results (see, for example, the discussion that follows on use of familiar content) and because finding ways to improve students' problem-solving performance is an educational goal, it is important to continue this line of research. Accordingly, this study investigated the relationship between contextual and conceptual wording and seventh-grade students' problem-solving achievement.

## **Review of Related Literature**

### **The Role of Problem Context and Language**

Problem context, the verbal or “worded” part of word problems, can influence student performance in solving word problems. For example, students' ability to engage in the problem context and meaningfully construct problem representations of relationships among key problem elements relates to their ability to solve word problems (Depaepe et al., 2010; Moore & Carlson, 2012; Voyer, 2011). Some evidence shows that unfamiliar contexts can be one source of student difficulty in solving word problems (Seifi, Haghverdi, & Azizmohamadi, 2012; Zahner, 2012). As a result, familiar contexts are often thought to improve student success in solving word problems by allowing problem solvers to make connections with prior knowledge and by engaging and motivating students to a greater degree (e.g., Jacobs & Ambrose, 2008/2009; Woodward et al., 2012). However, research is inconclusive regarding whether using familiar

content in word problem contexts improves students' ability to solve word problems. Some studies have not shown an achievement advantage by using familiar contexts for word problems (Huang, 2004; Inoue, 2008). For example, Inoue's (2008) study showed that undergraduate students performed better solving word problems that had fewer contextual constraints, thus allowing them to construct their own personal meaning for the problems.

Personalized problems, which incorporate individual student information (e.g. interests) into word problems, have been shown to improve student achievement in some cases but not others in solving problems, despite the fact that students often display favorable reactions to the problems (Awofala, 2011; Simsek & Cakir, 2009). In Awofala's (2011) study where students performed better on personalized problems, the lower-achieving students who solved these problems made the greatest gains among the 12- to 15-year-olds.

Problem context and linguistic complexity can influence the difficulty level of word problems for all students. However, the challenges of solving word problems are magnified for students whose first language is not the language of instruction; this is due to the importance of language proficiency in solving word problems (Barbu, 2010; Hoffert, 2009; Kempert, Saalbach, & Hardy, 2011; Wilburne, Marinak, & Strickland, 2011). Wilburne et al. (2011) note that attempts at "real-world" contexts do not necessarily represent the real world for all students "since mathematical word problems require students to understand language, culture, the context of the problem, and the mathematics, the problems can be rather complex and often confusing for culturally and linguistically diverse (CLD) students" (p. 461). The authors note that problem context containing holidays unique to the United States and many other culturally specific terms and concepts increase word problem difficulty for CLD students (specifically, cultural and linguistic minorities).

## **The Influence of Problem Rewording**

Some researchers have reworded the linguistic or syntactic aspects of word problems in an attempt to improve clarity and thus comprehension. For example, Eric (2005) conducted research in which upper-elementary students solved problems reworded in four different ways. In this study, students showed better comprehension and achievement on some types of reworded problems. Greater improvement resulted when semantic structures involved rearranging problem events into chronological order and repositioning problem givens in a manner that made mental representation of the problem clearer. Rewording involving personalization (described above) or chunking (breaking sentences into shorter pieces) did not aid student performance.

Mahajne and Amit (2009) studied a different type of problem rewording. They reworded word problems in a more colloquial manner that they considered to be closer to students' everyday language and investigated fifth- and sixth-grade students' achievement and attitudes in solving the problems. Students who received instruction and were tested using reworded problems showed significantly better achievement than the control groups. Fifth-grade students raised their achievement more than the sixth-grade participants. The study sample, especially the boys, also showed a significantly more positive attitude toward word problems. Improvements in both achievement and attitudes were stronger for students who were at a lower mathematics level and whose parents had a lower education level.

In another study that found differential influence of problem rewording, Samelson (2009) found rewording assisted first graders with low normal language skill but not those with language impairment. Finally, Vicente, Orrantia, and Verschaffel's (2007) research with third through fifth graders showed that students were more successful solving problems involving conceptual rewording (semantic relations made more explicit) than situational rewording (enhanced problem context), especially for younger students and with regard to more difficult problems. Some researchers thus consider rewording mathematical word problems one way to improve student problem-solving performance (e.g., Eric, 2005),

although research findings show that rewording has involved many different types of study design and has yielded mixed results.

## **Contextual and Conceptual Rewording**

In *contextual rewording* (also known as situational rewording), the context—or non-mathematical content—of a problem is altered. Changing the “storied” context may help give meaning to the mathematical content in a problem and is likely to influence, in particular, the problem-solving stages of understanding the problem and planning its solution (Kulm, 1984). As noted earlier, use of personalized and familiar story content can influence student problem-solving performance. Further, concrete problems are easier than abstract problem contexts and factual contexts are easier than hypothetical contexts (Caldwell & Goldin, 1987), indicating that differing levels of abstraction seem to influence problem-solving performance.

The following example illustrates a problem context that might be unfamiliar to seventh graders: “The weight of a beef pack is 6 kg. The weight of a hamburger pack is  $\frac{1}{3}$  of a beef pack. The weight of a piece of fish is  $\frac{1}{2}$  of a hamburger pack. What is the weight of a piece of fish?” A problem context revised to make it more personally familiar, while maintaining problem structure, is: “Ali bought a pack of 6 pencils. The number of erasers is  $\frac{1}{3}$  of the pencils. The number of pencil sharpeners is  $\frac{1}{2}$  of the erasers. How many pencil sharpeners does Ali have?”

In *conceptual rewording* of word problems, textual modifications make semantic relations among key problem elements (e.g., given and unknown sets) more explicit and transparent by, for example, adding or rearranging text (Eric, 2005; Vicente et al., 2007). The following word problem is an example of conceptual rewording:

Original problem: Peter had 37 meters of cable. He bought A meters of cable more. He used B meters of cable and ended up

with 11 meters of cable. How many meters of cable did he buy/use?

Conceptually reworded problem: Peter had 37 meters of cable. He bought A meters of cable more *and joined them with those that he had*. From the resulting total meters of cable, he used B meters and ended up with 11 meters of cable. How many meters of cable did he buy/use?

### **Student Errors in Solving Word Problems**

Knifong and Holtan (1976) classified word problem errors into "clerical and computational errors" and "other errors" (respectively, first type and second type errors). The "clerical and computational" category includes copying and calculation errors, and "other errors" involve use of wrong procedures or operations, no response, and incorrect responses offering no insight into the nature of the error.

### **Research Rationale and Purpose**

Research on rewording word problems has focused on a variety of types of rewordings, and the findings of these investigations—even those that address the same types of rewordings (e.g., problem personalization or familiarity)—have been mixed. Further, limited research has addressed differences in the types of errors students make in solving problems with different types of wording. Therefore, further research is needed on problem rewording in the field's continuing efforts to improve student problem-solving performance.

The purpose of this study was to investigate the influence of contextual, conceptual, and combined contextual/conceptual rewording of mathematical word problems on students' mathematical problem-solving performance. Specifically, the following questions guided the research: Does contextual, conceptual, and/or combined contextual/conceptual rewording of mathematical word problems influence seventh-grade students' achievement in solving word problems? Are there differences among these three categories of rewordings in student errors made?

## Method

The research methodology employed for this study was quantitative, specifically, a pretest-posttest design involving word problem solving with analysis of scores involving errors made. The word problems the experimental groups of students solved in the posttest were reworded to make the problem representation more explicit and thus easier to translate to a solution method through clearer identification of key elements and their interrelations. The independent variables are three types of problem rewording (contextual, conceptual, and combined contextual/conceptual), and the dependent variable is student scores for errors made in solving the word problems.

## Instrumentation

The first author developed four paper-and-pencil mathematics tests to serve as data collection instruments. Test  $T_0$  is the original (unmodified) test used as the pretest for all groups and for the control group's posttest. Tests  $T_1$ ,  $T_2$ , and  $T_3$  are reworded test versions matching the three independent-variable categories indicated previously.

Test  $T_0$  consists of six two-step arithmetic word problems taken from the standard seventh-grade mathematics textbook used in the country of Iran. The problems were selected with two characteristics in mind. One was that the problem context had to be considered unfamiliar or minimally familiar to participants. A second requirement was that the problem text offered little description to contextualize key problem elements or their interrelations.

All potential word problems were first assembled into a list. From these, five mathematics head teachers from the province in which the study site was located, none of whom served as staff in the school where the research was conducted, selected six problems meeting the two criteria noted above using the Delphi method (Fish & Busby, 2005). The participating teachers each had at least 15 years of teaching experience and held a certificate equivalent to a master's degree in mathematics education. The Delphi method involves

two or more rounds of action where experts anonymously provide responses to structured questions. They then receive summarized group decisions, anonymously revise their original input if they so choose, and repeat this process until the group reaches agreement.

Test  $T_1$  was created by rewording Test  $T_0$  problems to make the problem context more familiar. For Test  $T_2$ ,  $T_0$  problem statements were reorganized and reworded to make semantic relations between the given and unknown sets explicit. Test  $T_3$  was a combination of the two types of rewording used in  $T_1$  and  $T_2$ . The same five head teachers used a process similar to that used for developing the  $T_0$  test to develop the three reworded tests.

The following sample problems illustrate the types of problems used for the four test versions:

**Test  $T_0$  (Original Problem)**

The temperature in Ardabil is 10 degrees lower than Tehran. The temperature in Tehran is 15 degrees C. The temperature in Arak is 4 degrees more than Ardabil. What is the temperature in Arak?

**Test  $T_1$  (Contextual Rewording)**

Ali's score in mathematics is 10 points lower than in literature. Ali's score in literature is 15. Ali's score in geography is 4 points more than in mathematics. What is Ali's score in geography?

**Test  $T_2$  (Conceptual Rewording)**

The temperature of three cities, Tehran, Ardabil, and Arak, is expressed as: The temperature in Tehran is 15 degrees C. The temperature in Ardabil is 10 degrees lower than Tehran. The temperature in Arak is 4 degrees more than Ardabil. What is the temperature in Arak?

**Test  $T_3$  (Combined Contextual/Conceptual Rewording)**

Ali's scores in mathematics, literature, and geography are expressed as: Ali's score in literature is 15. Ali's score in mathematics is 10 points lower than in literature. Ali's



score in geography is 4 points more than in mathematics. What is Ali's score in geography?

## **Participants and Data Collection Procedures**

The study participants were 80 seventh-grade boys from three randomly selected classrooms in a single-sex school in Arak, Iran. The students' average age was approximately 12. The students were stratified into achievement categories based on their average scores in mathematics class. Students within each of these categories were then randomly assigned to one of the four word problem groups (one control and three reworded) in groups of 20 students each. The types of problems the students were used to solving prior to the study were real-world problems relevant to their lives.

All participants completed the pretest ( $T_0$ ) and later one of four posttests:  $T_0$  for the control group and  $T_1$ ,  $T_2$ , or  $T_3$  (tests involving different types of reworded problems) for the experimental groups. The control group took the same pretest and posttest in order to achieve confidence about homogeneity of the experimental groups and to determine whether or not repeating the test would enhance the control group's performance (i.e., reduce errors) to serve as a point of comparison for the experimental groups' results.

Across data collection forms, the order of problem presentation was randomly mixed. All tests were administered during regular school hours, the pretest in the morning and the posttest during the afternoon of the same day. Administering the pretest and posttest on the same day reduced potential intervening variables that might be introduced with a greater time gap and thus call into question obtained results, as in problem-solving instruction that might help participants develop greater problem-solving skill. Use of a control group further ensured trustworthy results in relation to the same-day test administrations due to rewording effects being investigated on experimental groups that were compared to a control group tested under the same conditions. Additionally, this research focused on the effects of different types of word problems on student performance (reduction of errors) rather than changes

within the learner, which would require greater time to gauge learner development and the stability of that development. Thus, no intervention (e.g., instruction) was introduced between the pretest and posttest. Finally, the posttest for the experimental groups differed from the pretest, as described earlier, and thus was likely to be perceived differently by participants.

Students completed each test in less than 30 minutes. Although the students solved word problems daily from their textbooks, they did not have explicit instruction in mathematical problem solving prior to administration of the research instruments.

## **Data Analysis**

In order to gather more information about types of errors made, each problem was scored as follows: If a student made one or more “Clerical and Computational Errors,” he earned a score of 1 for that category. The same scoring applied to the “Other Errors” category. A problem solution that showed at least one error in each category was assigned a score of 2, and a response with no errors earned a score of 0. Thus, the score for each problem ranged from 0–2. The error types were those determined by Knifong and Holtan (1976): (1) clerical and computation errors, and (2) all other errors. The first, consisting of copying and calculation errors, is here considered to be more procedural and superficial. The second is construed to be more conceptual and thus integral to the deep meaning of the problems. For this category, students used incorrect or unintelligible strategies, or they made no attempt to solve the problem, which might reflect a lack of understanding given that all problems were deemed accessible to participants and sufficient time was provided to solve them. Thus, we considered error type to be more important than error frequency in each category in order to provide more meaningful and global information. However, future research in this area might be warranted to refine errors into more nuanced categories.

Pretest and posttest scores were each analyzed using a one-way ANOVA to compare the four test conditions (control plus three types of rewording). Further, a post-hoc Tukey's HSD test was performed on the posttest scores to determine whether significant differences existed between each paired combination (six pairs total) of the four test groups. A separate chi-square analysis was performed on scores for each of the two types of errors in the three categories of reworded problems to determine whether the types of errors made in different types of reworded problems differed significantly.

## Results

Mean scores for pretest and posttest errors for the four test-condition categories are shown in Table 1. The results show that rewording in general (contextual, conceptual, and their combination) reduces student errors more than that of a control group comprised of students who solved problems that were not reworded. An analysis of variance (ANOVA) shows that the pretest-to-posttest scores across problems were significantly different (see Table 2).

Table 1  
*Pretest and Posttest Means and Standard Deviations for Word Problem Errors*

Test Condition	Pretest			Posttest	
	N	M	SD	M	SD
Control Group	20	5.50	1.43	4.95	1.32
Contextual Rewording	20	5.35	1.42	2.25	0.85
Conceptual Rewording	20	5.80	1.40	2.75	1.21
Combined Contextual/Conceptual Rewording	20	6.10	1.65	2.15	1.04

*Note.* Each test consists of six problems. Each problem was scored 0-2 for type of errors made (0 for none, 1 for at least one error in either the "clerical and computational errors" or "other errors" category, or 2 for at least one error in each category). Each participant's test score may range from 0-12.

Table 2  
*One-Way Analysis of Variance for Posttest Scores*

Source	SS	df	MS	F	Sig.
Between Groups	102.950	3	34.317	27.453	<.001
Within Groups	95.000	76	1.250		
Total	197.950	79			

A post-hoc Tukey's HSD test compared each possible pair of posttest scores (see Table 3). This test showed that the mean difference in scores between the control group and each of the three experimental groups (contextual, conceptual, and their combination) differed significantly ( $p < 0.001$ ), indicating that all three types of reworded problems resulted in significantly less problem-solving error compared with the control group. (Note that a one-way ANOVA performed on the pretest scores of the four test groups showed no significant differences.) However, paired comparisons among the three types of problem rewordings showed no significance differences in their effect on problem scores, meaning that no one type of reworded problem demonstrated an advantage over the others in reducing student errors.

Table 3  
*Post-hoc Tukey: Comparison of Posttest-Error Scores*

Pairs Compared	AB	AC	AD	BC	BD	CD
Mean Difference	.50	.10	2.7	.60	2.2	2.8
$p$	.495	.992	<.001	.332	<.001	<.001

*Note.* A = contextual rewording; B = conceptual rewording; C = combined contextual/conceptual rewording; D = control group.

Data presented in Table 4, which were analyzed using a chi-square test, did not differ significantly in terms of reduction of the first type of student error, clerical and computational errors ( $\chi^2 = 3.73$ ,  $p > 0.05$ ). However, data analyzed for the second type of student error (see Table 5) show that problem rewording significantly reduced other types of errors ( $\chi^2 = 137.1$ ,  $p < 0.05$ ). Thus, rewording problems influenced students

to make significantly fewer errors coded as wrong procedures or operations, no response, or incorrect responses offering no insight into the nature of the error, but rewording problems did not significantly reduce clerical and computational errors (e.g., copying and calculation errors).

Table 4  
*Contingency Table for "Clerical and Computational Errors"*

Experimental Group	Clerical and Computational Errors			
	Pretest Score	Posttest Score	$\chi^2$	<i>p</i>
Contextual Rewording	46	33	2.14	0.64
Conceptual Rewording	50	45	0.263	0.61
Combined Contextual/Conceptual Rewording	48	35	2.03	0.15

*Note.* Each experimental group completed the same pretest but a different posttest based on the indicated type of rewording. Each group consisted of 20 participants who completed six items per test, each of which was scored 1 if it contained one or more clerical/computational errors. Thus, each pretest or posttest score falls within a range of 0-120.

Table 5  
*Contingency Table for "Other Errors"*

Experimental Group	Other Errors			
	Pretest Score	Posttest Score	$\chi^2$	<i>p</i>
Contextual Rewording	61	7	42.88	<.01
Conceptual Rewording	66	10	41.26	<.01
Combined Contextual/Conceptual Rewording	74	8	53.12	<.01

*Note.* Each experimental group completed the same pretest but a different posttest based on the indicated type of rewording. Each group consisted of 20 participants who completed six items per test, each of which was scored 1 if it contained one or more clerical/computational errors. Thus, each pretest or posttest score falls within a range of 0-120.

## Discussion

This study provides data on the influence of three types of problem rewording (contextual, conceptual, and combined contextual/conceptual) on decreasing student errors in solving

word problems. The results show that rewording word problems, in general, can favorably influence student performance. This is most likely due to clarifying and extending information in a way that helps students to better conceptualize the problems, which is a vital step in choosing appropriate solution methods. Although research is somewhat mixed on the effect of rewording problems in ways that enhance contextual information, linguistic difficulty, and semantic relations, as detailed earlier in this article, the findings reported here support research yielding positive results from such problem rewordings (Awofala, 2011; Mahajne & Amit, 2009; Vicente et al., 2007).

No significant differences appear among the three rewording types, indicating that all types were important in helping students engage in the problems in ways that allowed them to be more successful in solving them. This means that using familiar contexts and stating problems in ways that make key problem elements and their relationships more obvious are useful in creating more “accessible” word problems. Of particular interest is the fact that the problem rewordings used in this study significantly influenced better problem solving in relation to Knifong and Holtan’s (1976) “other errors” category but not their “clerical and computational errors” category. This makes sense in that the “other errors” involve more effortful thinking and action (e.g., providing a meaningful response, as in choosing an appropriate solution method), whereas the “clerical and computational errors” category consists of superficial errors in copying or calculation.

We conclude that problem rewording is particularly important for deeper problem-solving efforts, such as conceptualizing problem structure in a meaningful way that facilitates choosing an appropriate course of action. This resonates with other research findings where more experienced or higher-achieving students seem to be less dependent on problem enhancement for successful problem-solving performance. For example, in Awofala’s (2011) study, low-achieving adolescents made greater gains than higher-achieving peers in solving personalized problems. Similarly, Mahajne and Amit’s (2009) and Vicente et al.’s (2007) studies

involving problem rewording show that younger students responded more favorably than older students in terms of problem-solving success. These two studies also found that problem rewording especially aided the problem-solving efforts of lower-achieving students and student attempts to solve problems that are more difficult. These collective findings might explain why undergraduate students in Inoue's (2008) study performed better when solving word problems with fewer contextual constraints. Because undergraduate students have greater academic experience and have demonstrated some degree of scholarly success, they might be less reliant on explicit information to help them solve mathematical word problems. By that point in their schooling, they have become quite familiar with standard word problem forms and might be hindered by constraints that limit their own intellectual input into a task.

It might be the case that younger, less experienced, and lower-achieving students benefit most from rewording problems in a way that provides more fully developed and explicit information. This suggests scaffolding problems by providing greater and clearer detail for younger and struggling students and for all students when solving more challenging problems. The additional structure can be gradually reduced over time. This is a matter of matching problem types to student need in a way that adapts to and supports students as their problem-solving abilities and experiences improve. It does not mean locking students into more "forgiving" or prescriptive problems without increasing challenge over time, because it is important to note that better performance on reworded problems does not equate with improved problem-solving ability. Finding ways to help students make sense of problems without "giving away" decisions that need to be made by the problem solver (namely, choosing a solution method and determining if an answer makes sense) is a reasonable consideration for improving problem-solving teaching and learning. The research reported in this study provides additional evidence to the field that supports this contention.

## Potential Limitations and Further Research

The original problems used in this study were selected from Iranian mathematics textbooks. Cultural and linguistic contexts influence the type of problem structure and contextual meanings that characterize problems. These factors would have to be considered when conducting similar research, because the word problems used would need to be familiar types within a given cultural context.

All participants in this study were male. In Mahajne and Amit's (2009) research described earlier, boys showed a more favorable attitudinal response to problems reworded in more colloquial language than did girls. This stronger response could favorably influence problem-solving performance. Females display more negative attitudes toward mathematics than males and have greater difficulty performing more complex mathematics (e.g., problem solving) than less complex mathematics (e.g., computation) when compared to males (Gibbs, 2010; Nosek & Smyth, 2011). It is important to determine how the findings reported here might compare to those obtained for females.

In this research, students received no problem-solving instruction before participating in the study. It might be worthwhile to determine whether students who participated in word problem instruction beforehand would respond similarly to the reworded problems used in this research.

## References

- Ahmad, A., Salim, S. S., & Zainuddin, R. (2008). A cognitive tool to support mathematical communication in fraction word problem solving. *WSEAS Transactions on Computers*, 4(7), 228–236.
- Awofala, A. O. A. (2011). Effect of personalized, computer-based instruction on students' achievement in solving two-step word problems. *International Journal of Mathematics Trends and Technology*, 2(2). Retrieved February 8, 2013, from [www.ijmtjournal.org/Volume-2/issue-2/IJMTT-V2I2P502.pdf](http://www.ijmtjournal.org/Volume-2/issue-2/IJMTT-V2I2P502.pdf)
- Barbu, O. C. (2010). Effects of linguistic complexity and math difficulty on word problem solving by English learners. *International Journal of Education*, 2(2), 1–19.



- Cai, J., & Lester, F. (2010). *Why is teaching with problem solving important to student learning?* [Research Brief]. Reston, VA: National Council of Teachers of Mathematics.
- Caldwell, J. H., & Goldin, G. A. (1987). Variable affecting word problem difficulty in secondary school. *Journal for Research in Mathematics Education*, 18(3), 187–196.
- Depaeppe, F., De Corte, E., & Verschaffel, L. (2010). Teachers' approaches towards word problem solving: Elaborating or restricting the problem context. *Teaching and Teacher Education*, 26(2), 152–160.
- Eric, C. C. M. (2005.) Language proficiency and rewording of semantic structures in P5 pupils' mathematical word problem solving. *The Mathematics Educator*, 9(1), 84–99.
- Fish, L. S., & Busby, D. M. (2005). The Delphi Method. In D. H. Sprenkle & F. P. Piercy (Eds), *Research methods in family therapy* (2<sup>nd</sup> ed.; pp. 238–253). New York: Guilford.
- Gibbs, B. G. (2010). Reversing fortunes or content change? Gender gaps in math-related skill throughout childhood. *Social Science Research*, 39(4), 540–569.
- Hoffert, S. B. (2009). Mathematics: The universal language? *Mathematics Teacher*, 103(2), 130–139.
- Huang, H-M. E. (2004). The impact of context on children's performance in solving everyday mathematical problems with real-world settings. *Journal of Research in Childhood Education*, 18(4), 278–292.
- Jacobs, V. R. & Ambrose, R. C. (2008/2009). Making the most of story problems. *Teaching Children Mathematics*, 15(5), 260–266.
- Kempert, S., Saalbach, H., & Hardy, I. (2011). Cognitive Benefits and costs of bilingualism in elementary school students: The case of mathematical word problems. *Journal of Educational Psychology*, 103(3), 547–561.
- Knifong, J. D., & Holtan, B. (1976). An analysis of children's written solutions to word problems. *Journal of Research in Mathematics Education*, 7(2), 106–121.
- Kulm, G. (1984). The classification of problem-solving research variables. In G. A. Goldin & C. E. McClintock (Eds.), *Task variables in mathematical problem solving* (pp. 1–21). Philadelphia: The Franklin Institute Press.
- Mahajne, A., & Amit, M. (2009). The effect of rephrasing word problems on the achievements of Arab students in mathematics. In L. Paditz & A. Rogerson (Eds.), *Proceedings of the Tenth International Conference "Models in Developing Mathematics Education"* (pp. 383–385). Dresden, Germany: The University of Applied Sciences.

- Moore, K. C., & Carlson, M. P. (2012). Students' images of problem contexts when solving applied problems. *Journal of Mathematical Behavior*, 31(1), 48–59.
- Nosek, B. A., & Smyth, B. A. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Education Research Journal*, 48(5), 1125–1156.
- Samelson, V. M. (2009). *The influence of rewording and gesture scaffolds on the ability of first graders with low language skill to solve arithmetic word problems*. (Unpublished doctoral dissertation). University of Iowa, Iowa City, Iowa.
- Seifi, M., Haghverdi, M., & Azizmohamadi, F. (2012). Recognition of students' difficulties in solving mathematical word problems from the viewpoint of teachers. *Journal of Basic and Applied Scientific Research*, 2(3), 2923–2928.
- Simsek, N., & Cakir, O. (2009). Personalization on students' achievement and gender factor in mathematics education. *International Journal of Human and Social Sciences*, 4(4), 278–282.
- Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology*, 77(4), 829–848.
- Voyer, D. (2011). Performance in mathematical problem solving as a function of comprehension and arithmetic skills. *International Journal of Science and Mathematics Education*, 9(5), 1073–1092.
- 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(2), 174–195.
- Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8: A practice guide* (NCEE 2012-4055). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved February 8, 2013, from [http://ies.ed.gov/ncee/wwc/publications\\_reviews.aspx#pubsearch/](http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch/)
- Zahner, W. (2012). “Nobody can sit there”: Two perspectives on how mathematics problems in context mediate group problem solving discussions. *Journal of Research in Mathematics Education*, 1(2), 105–135.