Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education) 39: 284-295 [2010]

EFFECT OF CAUSAL STORIES IN SOLVING MATHEMATICAL STORY PROBLEMS

SÖZEL MATEMATİKSEL PROBLEMLERİN ÇÖZÜMÜNDE NEDENSEL ÖYKÜLERİN ETKİSİ

Glenn Gordon SMITH[**], Helen GERRETSON[***], Sinan OLKUN[****], Jorma JOUTSENLAHTI[*****]

ABSTRACT: This study investigated whether infusing *causal* story elements into mathematical word problems improves student performance. In one experiment in the USA and a second in USA, Finland and Turkey, undergraduate elementary education majors worked word problems in three formats: 1) standard (minimal verbiage), 2) potential causation (causal and mathematical content overlap), and 3) climax resolution (causal and mathematical content combined in a way in which story outcome is discernable). Causal story elements in word problems, written in the USA, improved performance in USA and Finish students, but not Turkey, on word problems with some spatial content. Based on the finding that infusing causal stories played out differently in different cultures we concluded that situation models might be at least as primary as schemas in solving word problems.

Keywords: word problems, schema models, situation models

ÖZET: Bu çalışmada sözel problemler içerisine yerleştirilen nedensel öykülerin öğrenci performansını etkileyip etkilemediği araştırılmıştır. İlki Amerika'da, ikincisi ise Amerika, Türkiye ve Finlandiya'da eş zamanlı olarak yapılan deneysel çalışmalarda sınıf öğretmeni adayları 3 farklı tipte 1) standart (minimum sözel matematiksel bilgi), 2) nedensel (hem nedensel hem de sözel matematiksel bilgi içeren, 3) olası sonuç da içeren (problemin olası etkileri ve sonuçları görülen bir şekilde) hazırlanmış sözel matematiksel problemleri çözmeye çalışmışlardır. Nedensel öykü unsurlarının eklenmesi Amerikalı ve Finli öğrencilerin uzamsal içerikli problemlerdeki performansını artırırken Türk öğrencilerde herhangi bir etki yapmamıştır. Nedensel öykü unsurlarının eklenmesinin farklı kültürlerde farklı etki yapabileceği bulgusundan hareketle sözel problemlerin çözümünde durumsal modellerin en az şema modelleri kadar öncelikli olabileceği sonucuna varılmıştır.

Anahtar sözcükler: sözel problemler, şema modelleri, durumsal modeller.

1. INTRODUCTION

There is a current debate about whether solving mathematical word problems primarily involves schemas (invariant recipes for problem structures) (Kintsch & Greeno, 1985; Riley, Greeno, & Heller, 1983; Devidal, Fayol, & Barrouillet, 1997) or situation models (episodic models of the unique elements of the story) (Thevenot, Devidal, Barrouillet, & Fayol, 2007; Coquin-Viennot & Moreau, 2003). This question has important implications for how to best approach word problems. The current authors suggest a new method for addressing the schema versus situation model conundrum. Drawing from theories that readers create *situation models* while reading text narratives (van Dijk & Kintsch, 1983; Zwaan, Langston, & Graesser, 1995), the authors investigated whether combining *causal* elements with *spatial* elements of word problems boosted students' performance—i.e., does putting the story back into "story problems" improve performance on word problems?

1.1. Theoretical Frameworks

Textbooks and other curricular materials used in teaching mathematics are a contributing factor to poor performance on word problems. Wyndhamn & Säljö (1997) observe that word problems that appear in traditional mathematics textbooks usually represent re-contextualized forms of de-contextualized descriptions of everyday life situations that serve a specific purpose: to embed the numeric information in a story as an exercise for specific types of mathematical learning, such as addition or multiplication. Further, Hiebert et al. (1996) posit that solving such word problems cannot

** Assist. Prof. Dr., University of South Florida, USA, glenns@usf.edu
*** Assist. Prof. Dr., University of South Florida, USA, hpg@coedu.usf.edu.
**** Prof. Dr., Ankara University, Faculty of Educational Sciences, Turkey, olkun@education.ankara.edu.tr
***** Assist. Prof. Dr., University of Tampere, Finland, Jorma.Joutsenlahti@uta.fi

prepare students for everyday life, since students are not able to retrieve the needed specific domain-related (mathematical) knowledge. Various researchers (Silver, Shapiro, & Deutsch, 1993; Boaler, 2000; Schoenfeld, 1991; Verschaffel, Greer, & De Corte, 2000) attribute students' difficulties with word problems to students' lack of sense making or students' suspension of sense making when working with word problems, particularly in connecting computation to real life situations. Commenting on students' difficulty to apply common-sense knowledge when solving word problems, Verschaffel, De Corte and Vierstraete (1999) point out that "extensive experience with traditional arithmetic word problems induces in pupils a strong tendency to approach word problems in a mindless, superficial, routine-based way in their attempts to identify the correct arithmetic operation needed to solve a word problem" (p.265). Thus, the practice of word problem solving in school mathematics hardly matches the idea of mathematical modeling and *mathematizing*, which is the structuring of real life scenarios by mathematical means (Freudenthal, 1991). When students work with traditional word problems they tend not to use heuristics and mathematical strategies, but rather use mechanical and mindless solutions (Greer, 1997). A common behavior finds students looking for key words and employing direct translation strategies to solve a problem (Schoenfeld, 1992; Aydoğdu, Akbaba-Altun, & Olkun, 2004). All this suggests that there is much room for improvement in word problems, both in terms of improving student performance, but also in terms of writing better word problems. However there is still some confusion about the psychology of how students approach word problems.

There are two competing theories on the solving of word problems. One suggests that students solve word problems by plugging the particulars of a word problem into a schema, while the other suggests that students create a situation model of the story. The schema-based explanation of word problems suggests that after students successfully solve a number of similar types of word problems, they later access that schema from long term memory (LTM) when encountering another problem of the same type, using it to compute a solution (Rumelhart, 1980; Schank, 1975). The schema approach suggests that students create a "problem model" containing the minimal amount of information necessary to solve the problem.

The corollary of the problem model approach is that any extra elements included in word problems, above and beyond the minimal information needed to solve the problem, may exert an *extraneous cognitive load* (Sweller, 1994; Paas, Tuovinen, Tabbers, & Van Gernen, 2003) or redundancy (Chandler & Sweller, 1996), reducing performance. Cognitive load theory suggests that any type of problem has an *intrinsic cognitive load*, those minimal elements and the interactivity between those elements that the learner must hold in short term memory to solve the problem (Paas et al., 2003). Introducing any extraneous elements into the problem forces the learner to hold extra information in short term memory, potentially straining the learner's capacity to solve the problem if short term memory capacity is exceeded. For example, providing fifth graders with illustrations to go along with word problems does not improve performance and may actually hinder performance if learners must parse some information from the illustration and the rest of the information from the words (Berends & van Lieshout, 2009). Cognitive load factor may be important in word problems if learners primarily employ a schema approach creating a "problem model" of the vital elements of a word problem.

In contrast, the situation model explanation to solving word problems suggests that students approach word problems as stories, as they would approach other text narratives, i.e., by first creating a situation model of the unique elements of that story and then using any schemas secondarily. The situation model theory of how readers read text narratives may suggest that further contextualizing mathematical word problems could improve or at least not reduce student performance and learning. As people read text narratives, they read both at a surface level, parsing words and sentences, and at a deeper level, creating a cognitive situation model of the story and updating that situation model as events unfold. Readers' situation models include five dimensions (Zwaan et al., 1995): protagonist (who the main characters are), goal (of the characters), causation (how events cause changes in the situation), time/temporal (flashbacks and flash-forwards, etc), and space/spatial (where in the setting events take place). Of these dimensions, readers handle the spatial dimension the most poorly, typically opting out of creating and/or maintaining a spatial situation model of a story setting (Hakala, 1999; O'Brien & Albrect, 1992; Wilson, Rinck, McNamara, Bower, & Morrow, 1993).

Readers consistently and spontaneously track causal (Long, Golding, & Graesser, 1992; Trabasso & Magliano, 1996) and temporal (Rinck, Hahnel, & Becker, 2001; Zwaan, 1996) relations in text

narratives, but do not spontaneously track spatial relations (de Vega, 1995; Levine & Klin, 2001; Zwaan, Radvansky, Hilliard, & Curiel, 1998; Zwaan & van Oostendorp, 1993). This may be because creating a spatial situation model often requires visual mental imagery, which requires a decision and willful effort (Hasher & Zacks, 1979), and ongoing attention to maintain it in visual short term memory (Awh & Jonides, 2001; Shah & Mijake, 1996).

Combining *causal* with *spatial* elements increases reader monitoring of spatial elements (Jahn, 2004); readers mentally model the spatial situation in a predator-prey relation more frequently than in a prey-prey relation. The spatial relation in "*Two zebras move away from a shrub and a lion trots towards them*" is more likely to be recalled than in "*Two zebras move away from a shrub and an antelope trots towards them*" (Jahn, 2004, p. 973). The causal relation in this case is *potential*. The lion may or may not catch the zebras. However, perhaps for evolutionary reasons, the mere potential of the lion catching the zebras is enough to encourage reader to monitor the spatial situation.

The lead author of the current paper hypothesizes a second form of causation, i.e., causation resolution, or climax resolution, i.e., "*Two zebras move away from a shrub and a lion trots towards them, snapping the neck of one of them.*" There is a qualitative difference between potential causation (events that have a propensity to happen), and causation resolution (events with a propensity to happen that actually play out). This is the familiar conflict and resolution in stories. Stories have certain expected parts, a story grammar, that readers/listeners/viewers expect (Mandler & Johnson, 1977; Stein & Glenn, 1979). The *setting* describes the initial situation of the story and introduces the protagonist. The *beginning* produces a conflict from an initiating event, resulting in a *reaction* by the protagonist, who initiates some *attempts* or actions aimed at resolving the conflict, resulting in an *outcome*, followed by the *ending* of the story. Readers tend to remember events in the story that relate to these categories. For example, in Mandler and Johnson's seminal work (1977), after hearing recordings of stories, first graders, fourth graders, and adults all recalled far more from settings, beginnings, and outcomes than from any other parts of the stories. Later studies have reproduced and extended these results supporting story grammar (van den Broek, Lorch, & Thurlow, 1996). Thus one might expect that readers creating situation models from text narratives might react differently to potential causation (conflict) and causation resolution (climax resolution). Both of these forms of causation may be tools for encouraging readers to monitor spatial information in text.

Since spatial and mathematical thinking are related (Wheatley, 1998; van Garderen & Montague, 2003; National Council of Teachers of Mathematics, 2000; Cruz, Febles & Díaz, 2000), combining *potential causal* elements with mathematical elements in word problems may boost monitoring of the mathematical situation in the story, particularly if the word problems have spatial content. The current authors further hypothesized that the climax resolution format would also result in better word problem performance than typical formulations of word problems.

If the schema explanation of word problems is valid for a class of word problems, one would predict that a minimalistic presentation of the problem, with little more than the mathematical elements, would result in more correct answers. However, if the situation model explanation of solving word problems is valid for a class of word problems, one would expect that providing additional meaningful contextual information to the word problem (in other words, making it a better story) might motivate students to construct a more detailed situation model, resulting in more correct answers. In all probability, different approaches are used by different people on different classes of word problems. But the power of causality may come to bear with spatially oriented word problems.

Because including causal elements in word problems is new, there are many unknown elements. We do not know if including causality in mathematical word problems has any effect one way or the other. If there is such an effect, we do not know the bounds of such an effect. Does it work with all word problems, or only those with spatial content? Is the effect maximized with non-routine word problems or routine word problems? Is the effect attenuated with certain ethnicities or age groups? Considering the number of unknowns, this first series of studies is exploratory.

1.2. Research question

In mathematical word problems with spatial content, how do the following formats compare for student performance: 1) standard (minimal verbiage), 2) potential causation (causal and mathematical

content overlap), and 3) climax resolution (causal and mathematical content combined in a way that is the outcome is computable by the reader)?

Hypothesis 1: In mathematical word problems with spatial content, potential causation format will result in better student performance than will standard format.

Hypothesis 2: In mathematical word problems with spatial content, conflict resolution format will result in better student performance than will standard format.

The independent variable was the type of problems. There are three word problem formats: standard, potential causation, and climax resolution. The dependent variables were correctness of answers.

2. METHODS OF STUDY ONE

2.1. Participants

The participants were 72 female undergraduate elementary education students enrolled in a mathematical methods course as a teacher preparation program requirement at a state university in the south east United States. The average age of the participants was unknown as the students participated anonymously in the experiment. Undergraduate elementary education students were considered an appropriate population for this study because: a) pre-service elementary students often struggle with word problems (Chapman, 2003; Contreras & Martinez-Cruz, 2003) and pre-service teachers, in general, often demonstrate the same misconceptions observed in their students (Lord & Holland, 1997; Rosales, Orrantia, Vicente, & Chamoso, 2008; Sunberg & Goodman, 2005), and b) in the near future, they will be teaching word problems to students.

2.2. Materials

The investigators developed three very simple word problems that also involved some spatial content, in three contextualized versions: standard, potential causation, and climax resolution. The standard format included almost the minimal amount of information needed to solve the problem. Standard format: "*At seven o'clock in the evening, Pete the Frog fell into a damp, slippery well. As Pete ascends the nine-foot well, each hour he climbs three feet up, but slides one foot back down again. Between which hours will Pete reach the top of the well?*"

The potential causation format included, in addition to the sentences in the standard version, several sentences which link the mathematical content to potential causation in the story. In the following, the additional potential causation sentences are bolded. Potential causation format: *"At seven o'clock in the evening, Pete the Frog fell into a damp, slippery well.* ***He immediately began to climb up the wall to escape. If he doesn't get to the top of the well before the temperature reaches freezing, Pete will die.*** *As Pete ascends the nine-foot well, each hour he climbs three feet up, but slides one foot back down again. Between which hours will Pete reach the top of the well?"*

In the climax resolution format, mathematical content is connected to potential causation in the story; additionally, enough information is included so that the student can use the mathematical content to compute the outcome of the story. In the following, the key additional climax resolution sentence is bolded. The addition of the phrase "by midnight" allows the reader to compute the outcome of the story. Climax resolution format: *"At seven o'clock in the evening, Pete the Frog fell into a damp, slippery well. He immediately began to climb up the wall to escape.* ***If he doesn't get to the top of the well by midnight, when the temperature reaches freezing, Pete will die.*** *As Pete ascends the nine-foot well, each hour he climbs three feet up, but slides one foot back down again. Between which hours will Pete reach the top of the well?"*

2.3. Procedure

During a one-time session, as a non-graded portion of their course, students convened in a computer lab and sat down at computers with a web browser already at the correct URL. Students logged onto the page with an experimenter-supplied user ID and password. Then, students were presented with three different word problems: one in standard, one in potential causation, and one in climax resolution format. The presentation of materials was counter-balanced in terms of problem, format, and order. Therefore, one third of the students read one particular word problem in standard

format, one third of students read that same problem in potential causation format, and one third read it in climax resolution format. Every student encountered all three of the word problems. The students read each word problem one sentence at a time. This arrangement was done in order to get the reading times for the sentences. Longer reading times are associated with more cognitive processing of the underlying meaning of the text (Zwaan et al., 1995). After reading each word problem one sentence at a time, students were presented with the full word problem as one paragraph, along with a text box to record their answer. They were also given scratch paper, collected as data, should they need to write or draw anything associated with the problem. They were instructed not to write anything down, nor compute anything, during the sentence-by-sentence presentation, but to solve the problem in the full paragraph presentation. Unfortunately the times for reading sentences turned out to be unusable as some students, despite directions not to do so, wrote down information during the sentence by sentence presentation of the word problems, resulting in unrealistically long reading times. Additionally, to control for attitudes about reading, students also took the Mikulecky Behavioral Reading Attitude Measure (Schaer, Ley, Neal, & Wright, 1988).

3. RESULTS OF STUDY ONE

Since the three word problems were not standardized, each of three word problems was effectively a separate between-subject experiment. Two of the investigators (one of whom is mathematics educator) graded the students' answers to the word problems, assigning a value of 1.0 to correct answers, 0.5 to partially correct (with some valid logic but not 100% correct), and 0.0 to totally incorrect answers with fallacious logic. The inter-rater agreement across all problems was 83%.

For the first word problem (the frog in the well word problem as previously described), the mean average correctness was 0.65. Students presented with the potential causation format performed best, followed by climax resolution, then, followed by standard format. Based on hypothesis 1, that students would perform better with the potential causation format than with the standard format, we conducted a planned comparison t-test. On an independent sample t-test comparing average fraction correct between participants who answered the frog problem in *potential causation* format ($M = .79$, $SD = .319$) versus those who answered it in *standard format* ($M = .49$, $SD = .455$), the difference was significant, $t(2, 48) = 2.7$, $p < .009$, $d = .77$.

Also, to address hypothesis 2, that students would perform better with the climax resolution format than with the standard format, we conducted another planned comparison t-test. On this independent samples t-test, comparing average fraction correct between participants who answered the frog problem in *climax resolution* format ($M = .66$, $SD = .405$) versus those who answered it in *standard format* ($M = .49$, $SD = .455$), the difference was not significant, $t(2, 46) = 1.34$, $p < .181$.

For the other two word problems, the overall mean average correctness (0.13 and 0.3) was too low to be considered reliable. There were also no significant differences between groups for these two word problems in planned t-tests. To the investigators who graded the word problems, it appeared that many students misinterpreted the problems. For example in one problem (the "newlywed problem"), students were asked to compute the husband and wife's relative commute distances to work in Manhattan distance (distance along city blocks), however some students were incorrectly trying to apply the Pythagorean theorem to compute distance along the diagonal as the crow flies.

Data for reading times for sentences shared by all three formats were not reliable, as there were many invalid data points because some participants did not follow the directions to avoid writing things down during the initial sentence-by-sentence presentation of the problems, still need to be removed. Therefore these data were not reported.

4. METHODS OF STUDY TWO

The investigators conducted a follow-up study that essentially replicated the first study except with the following changes:

a) The stories were refined to remove any possible ambiguities, as informed by the first study. For example, in the first study on the newlywed problem, some students who did not understand that the problem called for Manhattan distance (distance along city blocks) incorrectly applied the

Pythagorean Theorem to compute distance. In the revised version of the newlywed problem, investigators attempted to remove this ambiguity. Following are the three new versions of the newlywed word problem:

Standard:

The city Jen and Darren live in is designed on a grid of rectangular blocks. Darren's job site was 20 miles north and 70 miles east of Jen's. They move into an apartment 30 miles south and 45 miles east of Jen's job. How far of a drive (not as the crow flies, but along the streets on the grid) is their apartment from Jen and Darren's job respectively?

Potential causation:

Even though Jen and Darren loved each other, they both found it hard to put the needs of the other before their own. They want to find an apartment that is an equal drive from their jobs, otherwise Jen will think it unfair. The city Jen and Darren live in is designed on a grid of rectangular blocks. Darren's job site was 20 miles north and 70 miles east of Jen's. They move into an apartment 30 miles south and 45 miles east of Jen's job. How far of a drive (not as the crow flies, but along the streets on the grid) is their apartment from Jen and Darren's job respectively?

Climax resolution:

Even though Jen and Darren loved each other, they both found it hard to put the needs of the other before their own. They want to find an apartment that is an equal drive from their jobs, otherwise Jen will think it unfair **and ultimately will divorce Darren**. The city Jen and Darren live in is designed on a grid of rectangular blocks. Darren's job site was 20 miles north and 70 miles east of Jen's. They move into an apartment 30 miles south and 45 miles east of Jen's job. How far of a drive (not as the crow flies, but along the streets on the grid) is their apartment from Jen and Darren's job respectively?

b) The word problems were presented to the students in paper form. The investigators believed that the paper format was much more natural than the computer interface to the participants (undergraduate elementary education majors). Since, the reading time data in the first study was problematic, the investigators decided the paper format might involve fewer confounding variables.

c) In order to further explore the range of story types that the constructs of schema versus story situation model might apply to, a fourth word problem was added, which did not contain any spatial aspects. This fourth problem was written by the Turkish co-investigators. The other three problems were written by the American investigators. The total number of word problems was four.

d) In order to see how cultural and ethnic factors might affect these issues, the study was conducted with equivalent sample populations of undergraduate elementary education majors in three different countries (USA, Finland, and Turkey), with the stories and directions translated into their native languages. The number of participants from the USA, Finland and Turkey was 112, 60, and 69, respectively.

In each country, the answers to the word problems were graded independently by two experts. In the USA, Finland, and Turkey, the inter-rater reliabilities were .85, .94 and .88 respectively. The averages between the two scores given by the two raters were used for the analysis.

5. RESULTS OF STUDY TWO

With four word problems, each with three formats, in three different countries, the results are voluminous. Therefore, we present the overall results in summary and then highlight the significant results in detail. Given the potential for cultural differences between countries, data from each country were analyzed separately. Further, since the word problems were not calibrated to be equivalent, each word problem was also analyzed separately.

As in the first study, planned comparison t-tests were conducted for each problem with formats (standard, potential causation, and climax resolution) as the independent variable and correctness of answers as the dependent variable. Based on the hypotheses, investigators also performed planned

comparisons, using t-tests to compare students' results on the potential causation versus standard formats and on the climax resolution versus standard formats.

For the students in the USA, only one of the word problems produced significant differences, while the other three problems did not. On the newlywed problem, students performed better on the climax resolution format than on the standard format. A planned comparison t-test comparing correctness in the *climax resolution* format (M = .54, SD = .467) versus correctness in *standard format* (M = .34, SD = .334) indicated a significant difference, t (69) = 2.12, p < .038, d = .5. On the other three word problems, the differences were non-significant.

Similarly, for the students in Finland, only one word problem—also the newlywed problem—produced significant differences between formats. On the newlywed problem, students did best on the climax resolution format, second best on the potential causation format, and worst on the standard format. On the planned comparison t-test between performance of students working the *climax resolution* format (M = .99, SD = .057) versus performance of students working the *standard format* (M = .69, SD = .249), the difference was significant, t (38) = 5.3, p < .0001, d =1.9. On the t-test comparing the *potential causation* format (M = .86, SD = .275) versus the *standard format* (M = .69, SD = .249), the difference was also significant, t (39) = 2.1, p < .043, d = .65.

In the Finnish study on three of the four problems, students performed best on the climax resolution format. However, the differences were significant for only one of the problems. It is notable that the sample sizes were modest.

For the students in Turkey, in three of the four problems students did marginally better with the standard format. However, these differences were only significant on one of the problems for one of the t-tests. On the t-test comparing the *potential causation* format (M = .63, SD = .336) versus the *standard format* of the frog problem (M = .82, SD = .252), the difference was significant, t (44) = 2.1, p < 0.041, d = .65.

6. DISCUSSION and CONCLUSION

These two studies explore the idea of infusing story elements, specifically causation, into mathematical word problems. The intermittently significant results suggest that causation may improve students' performance in mathematical word problems in certain situations, but that the extent and bounds of the effects still need to be determined. At the very least, adding causation does not hurt student performance.

In the first study, the overall mean average correctness (0.13 and 0.3) was too low to be considered reliable for two of the three word problems. The average for the remaining item, the frog problem, was sufficiently high. Participants performed significantly best on the frog problem in the potential causation format and significantly worse with the standard format. The main significant results from second study were that, in both USA and Finland, but not in Turkey, participants did significantly better with the climax resolution format than they did with the standard format on one of the word problems, namely the newlywed problem, which was written in the USA.

The Turkish students did significantly better with the standard format on one problem, the frog problem (written in the USA), and non-significantly better with the standard format on two of the other problems (also written in the USA). On one problem written in Turkey, the school commute problem, which was non-spatial, the Turkish students did approximately equal with the different formats.

Thus, USA and Finish students behaved as one block, Turkey as another. The obvious explanation is that the American and Finnish cultures, predominantly Christian and with a higher standard of living, number 12^{th} and 15^{th} in worldwide rating of Human Development Index (Human Development reports, 2008; Human Development Index, 2008), are relatively more similar, while the Turkish culture, predominantly Islamic with a lower standard of living, rated number 84^{th} worldwide in Human Development Index (Human Development Reports, 2005), is relatively more different from the other two countries. Thus the causal story frameworks written in the USA were meaningful in the USA and Finland, but not in Turkey. For example, the causative element in the climax resolution format of the newlywed problem was the threat of divorce, but divorce rates are very different in these three countries: Turkey 6%, Finland 51.2 % and USA 54.8% (Divorce Rates Around the World, 2007) Thus, the use of divorce as a story context might be much more meaningful in USA and Finland than in Turkey.

Another explanation for this figure could be the fact that the Turkish students face more standard (sterile) word problems than realistic, real life problems during their education in schools and other private

tutoring sessions compared to their peers in the USA and Finland because of the standard, nationwide examinations. In these exams no extraneous information is given in the problems to make them more objective and standard. Only the information relevant to the solution of the problem is given.

In the second study students in both Finland and the USA did better on the climax resolution format for one problem. In the first study, the USA students did better on one problem with the potential causation format. Thus, the number of significant results modestly favored the climax resolution and potential causation formats over the standard format. These results suggest that for this population and these types of word problems, providing causative story context is more effective than paring down problems to almost minimal mathematical schema. Additional story context does not produce extraneous cognitive load but adds contextual meaning and the motivation to find the solution.

Why the frog problem produced significant results in the first study, but not in the second study, is an unresolved problem. However, since the problems in the second study were less ambiguous, the results of study two are considered more dependable.

In the second study, the major pattern across countries was that students in Finland and the USA did significantly better on the newlywed problem in the climax resolution format, than in standard format. This supports the situation model approach to word problems. However, it brings up an important inquiry: How is the newlywed problem different from the other problems? Why did the newlywed problem produce a result favoring the climax resolution format when the other three problems did not? There were several key differences between the newlywed problem and the other problems: a) it was more spatial than the others, requiring the drawing of a map-like diagram with a coordinate system to solve it. Participants often solved the other three problems without drawing any type of a spatial diagram. b) It was a non-routine problem, not amenable to common computational formulas or algorithms.

The researchers posit that because the newlywed problem was more spatial and non-routine, it took more work to solve it, and thus the additional causative story context provided additional motivation. It requires a decision and effort of will to conjure up visual mental images of shapes (Hasher & Zacks, 1979), and then ongoing attention to maintain them in the mind's eye (Awh & Jonides, 2001; Shah & Mijake, 1996). The results of this second study suggest that adding causative story context to word problems can improve student performance, but probably only in cases where the word problem is spatial and the mathematical schema is unfamiliar to students. However, these may be important learning situations for students.

Educationally, even though the mathematical content was identical in all formats across both studies, with some problems students solved word problems in potential causation and climax resolution format significantly better than in standard format. However, if situation models are more primary than schemas in word problems, then perhaps the current way of teaching word problems (as problem structures, motivated by the schema theory) is misguided.

This suggests that (at least for elementary education majors solving word problems) quality of story may be more important than minimizing cognitive load. Involving the mathematical content with causal elements of the story motivates students to correctly work word problems. Efforts to redesign word problems so that mathematics content overlaps with causal elements of the storyline might result in increased learning of word problems.

Currently, there is a movement towards making word problems culturally more relevant—for example, talking about the ocean when you are in Florida and the mountains when you are in Colorado. However, adding context should not be done superficially. Given that the current results suggest that students treat word problems fundamentally as stories, context should be added in a way that makes the mathematics content more fundamental to the story, hence making the mathematical elements part of the plot. Furthermore, since there is mounting evidence that situation models are at least as primary as schemas in solving word problems, perhaps some research should go into using situation model theory to articulate strategies for learning and teaching word problems. The current results also emphasize that word problems with story elements written in one culture may transfer well to similar cultures, but not to dissimilar cultures.

Principles of situation model theory—that causal-spatial overlap improves spatial dimension of the situation model while reading (Jahn, 2004)—seem to extend to an overlap between causal and mathematical content in word problems. These results are limited in scope and further studies need to be undertaken with more word problems to better delineate the types of word problems that could be improved with infusion of causal story elements. Further, it would be valuable to extend the current research paradigm to other populations such as elementary and middle school students to see how developmental differences in mathematics ability interact with story elements in word problems.

In this exploratory study of word problems, we conclude that infusing causation into word problems for undergraduate elementary education majors may improve student performance in non-routine word problems with spatial content. In cases of non-standard word problems with spatial content, if cultural differences are taken into consideration, it may improve student performance to put the story back into story problems.

KAYNAKLAR

Alwyn, O., & Wearne, D. (1996). Problem Solving as a Basis for Reform in Curriculum and Instruction: The Case of Mathematics, *Educational Researcher*, 25(4), 12-21.

Aydoğdu, T., Akbaba-Altun, S. & Olkun, S. (2004). İlköğretim 3, 4, ve 5. Sınıf Öğrencilerinin standart sözel problemlerde işlem seçimleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 27, 126-134.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences, 5*, 119-126.

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

Boaler, J. (2000). Introduction: Intricacies of knowledge, practice, and theory. In Boaler, J. (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 1-18). Mahwah. NJ: Lawrence Erlbaum Associates.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*, 151-170.

Chapman, O. (2003). Teachers' conceptions of mathematical word problems: A basis for growth. In N. A. Pateman, B. J. Dougherty, & J. T. Zillox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education* (vol. 2), (pp. 197-204). Honolulu: University of Hawaii.

Contreras, J.N., & Martinez-Cruz, A.M. (2003). Pre-service elementary teachers' solution processes to problematic addition and subtraction word problems involving ordinal numbers and their interpretations of solutions. In N. A. Pateman, B. J. Dougherty & J. T. Zillox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education* (vol. 2), (pp. 236-244). Honolulu: University of Hawaii.

Coquin-Viennot, D., & Moreau, S. (2003). Highlighting the role of the episodic situation model in the solving of arithmetical problems. *European Journal of Psychology of Education, 3*, 267-279.

Cruz, I., Febles, M., & Díaz, J. (2000). Kevin: A visualiser pupil. *For the Learning of Mathematics, 20*(2), 30-36.

de Vega, M. (1995). Backward updating of mental models during continuous reading of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 373-385.

Devidal, M., Fayol, M., & Barrouillet, P. (1997). Stratégies de lecture et résolution de problèmes arithmétiques. *L'Année Psychologique, 97*, 9-31.

Divorce Rates Around the World. (2007). Retrieved July 16, 2009, from http://www.darndivorce.com/divorce-rates-around-the-world/

Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht: Kluwer Academic Publishers.

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

Hakala, C. M. *(*1999). Accessibility of spatial information in a situation model. *Discourse Processes, 27*(3), 261-279.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 3*, 356-388.

Hiebert, J., Thomas P., Carpenter, E., Fennema, K., Fuson, P., Hanlie, M.,

Human Development Index. (2008). Retrieved July 16, 2009, from http://en.wikipedia.org/wiki/Human_Development_Index

Human Development Reports. (2008). Retrieved July 16, 2009, from http://hdr.undp.org/en/mediacentre/news/title,15493,en.html

Human Development Reports. (2005). Retrieved July 16, 2009, from http://hdrstats.undp.org/countries/country_fact_sheets/cty_fs_TUR.html

Jahn, G. (2004). Three turtles in danger: Spontaneous construction of causally relevant spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 969-987.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*(1), 109-129.

Levine, W. H., & Klin, C. M. (2001). Tracking of spatial information in narratives. *Memory & Cognition, 29*, 327-335.

Long, D. L., Golding, J. M., & Graesser, A. C. (1992). A test of the on-line status of goal-related inferences. *Journal of Memory and Language, 31*, 634-647.

Lord, T., & Holland, M. (1997). Preservice secondary education majors and visual-spatial perception: An important cognitive aptitude in the teaching of science and mathematics. *Journal of Science Teacher Education, 8*(10), 43-53.

Mandler, J. M, & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*, 111-151.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(4), 777-784.

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153-196). New York: Academic Press.

Rinck, M., Hahnel, A., & Becker, G. (2001). Using temporal information to construct, update and retrieve situation models of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 67-80.

Rosales, J., Orrantia, J., Vicente, S., & Chamoso, J. (2008). Studying mathematics problem-solving classrooms: A comparison between the discourse of in-service teachers and student teachers. *European Journal of Psychology of Education, 23*(3), 275-294.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theorical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education* (pp. 33-58). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Schaer, B. B., Ley, T. C., Neal, K. S., & Wright, J. P. (1988). Evaluation and validation of the Mikulecky behavioral reading attitude measure. *Educational and Psychological Measurement, 48*, 181-186.

Schank, R. C. (1975). *Conceptual information processing*. Amsterdam: North-Holland.

Schoenfeld, A. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*. (Ch. 15) New York: Simon & Schuster Macmillan.

Schoenfeld, A. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. In J. Voss, D. Perkins & J. Segal (Eds.), *Informal reasoning and education* (pp. 311-343). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shah, P., & Mijake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*, 4-27.

Silver, E., Shapiro, L., & 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*(2), 117-135.

Stein, N. L., & Glenn, C. C. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing: Advances in discourse processing* (pp. 53-120). Norwood, NJ: Ablex.

Sunberg, S., & Goodman, T. (2005). Incorporating spatial ability instruction in teacher preparation. *Mathematics Teaching in the Middle School, 11*(1), 28–34.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312.

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.

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes, 21*, 255-287.

van den Broek, P., Lorch, E. P., & Thurlow, R. (1996). Children's and adults' memory for television stories: The role of causal factors, story-grammar categories, and hierarchical level. *Child Development, 67*, 3010-3028.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

van Garderen, D., & Montague, M. (2003). Visual-spatial representation, mathematical problem solving, and students of varying abilities. *Learning Disabilities Research & Practice, 18*(4), 246-254.

Verschaffel, L., De Corte, E., & Vierstraete, H. (1999). Upper elementary school pupils' difficulties in modeling and solving nonstandard additive word problems involving ordinal numbers. *Journal for Research in Mathematics Education, 30*(3), 265-285.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets and Zietlinger.

Wheatley, G. W. (1998). Imagery and mathematics learning. *Focus on Learning Problems in Mathematics, 20*(2-3), 65-77.

Wilson, S. G., Rinck, M., McNamara, T. P., Bower, G. H., & Morrow, D. G. (1993). Mental models and narrative comprehension: Some qualifications. *Journal of Memory and Language, 32*, 141-154.

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(4), 361-382.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1196-1207.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing approach. *Psychological Science, 6*, 292-297.

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading, 2*, 199-220.

Zwaan, R. A., & van Oostendorp, H. (1993). Do readers construct spatial representations in naturalistic story comprehension? *Discourse Processes, 16*(1-2), 125-143.

Geniş Özet

Çocuklarda problem çözme becerilerinin geliştirilmesi matematik eğitiminin önemli amaçlarından birisidir. Bu beceriler ise çocukların daha basit durumlardan başlayarak giderek daha karmaşık durumlar içeren problemleri çözmeleri sağlanarak geliştirilmeye çalışılmaktadır. Problemler ise genellikle içinde sözel matematiksel durumlar içeren öykülerden oluşmaktadır. Sözel problemler olarak adlandırılan bu problemlerin çözümünü açıklayan rekabet içindeki iki kuramsal modelden birisi *şema*

modelidir. Bu model matematiksel sözel problemlerin çözümünün aynı tür problemlerin çözümünde kullanılan *şemalar* içerdiğini ileri sürmektedir. Durum modeli olarak adlandırılan diğer model ise insanların öykü metinlerini okurken bir yandan da öyküde betimlenen *durumun modelini* oluşturduklarını savunmaktadır. Sözel problemler içine uygun miktarda nedensel cümleler ekleyerek farklı yapısal özelliklerde problemler oluşturmak olanaklıdır. Öğrencilerin farklılaştırılan bu problemleri çözüm performansları farklı olursa bu iki modelden hangisini daha çok kullandıkları ortaya çıkabilecektir. Bu çalışmada sözel problemler içerisine yerleştirilen nedensel öykülerin öğrenci performansını nasıl etkilediği araştırılmıştır. Bulgular hangi modelin sözel problemlerin çözümünü açıklamada daha öncelikli olduğu konusunda bize fikir verecektir.

İlki Amerika'da, ikincisi ise Amerika, Türkiye ve Finlandiya'da eş zamanlı olarak yapılan deneysel çalışmalarda sınıf öğretmeni adayları 3 farklı tipte hazırlanmış sözel matematiksel problemleri çözmeye çalışmışlardır. Bunlar: 1. Standart, minimum sözel bilgi, 2. Nedensel, hem nedensel hem de matematiksel bilgi içeren, 3. Olası sonuç da içeren, problemin olası etkileri ve sonuçları problem içinde görülen şekilde hazırlanmış problemlerdir. İlk çalışma Amerikalı 72 öğrenci, ikinci çalışma ise 112 Amerikalı, 60 Finlandiyalı ve 69 Türk öğrenci olmak üzere toplam 312 öğrenci üzerinde uygulanmıştır. Her öğrenci 4 farklı problemin 3 değişik formatından birini çözmeye çalışmıştır. Böylece bir öğrenci aynı problemin farklı formatları ile karşılaştırılmamıştır. Kullanılan problemlerin zorluk düzeyi birbiriyle karşılaştırmalı olarak bilinmediğinden her bir problemin verisi ayrı olarak analiz edilmiştir. Her üç ülkede de aynı problemler aynı format ile yine aynı koşullarda kullanılmış ancak her ülkenin kendi dilinde sorulmuştur. Problemlerden bir tanesi diğerlerine göre daha çok uzamsal içerik içerecek şekilde hazırlanırken bir diğeri daha çok aritmetik içerikli olarak hazırlanmıştır. Araştırmada kullanılan dört problemden biri olan "bahçe" probleminin 3 farklı formatta yazılmış şekli aşağıda görülmektedir.

Standart: Harun'un babası, bir bahçe yapmak için evlerinin yanında, kırk dörde yetmiş altı metrelik küçük bir arsanın toprağını sürmüştür. Ertesi gün Harun'un küçük kardeşi, oyun oynarken, bir roketin kapsülünü bahçenin kenarına bir yere gömer. Harun ve kardeşi bir saat içinde bahçenin kenarları boyunca doğrusal olarak 40 metre uzunluğundaki bir yeri arayabilmektedirler. Bahçenin tüm çevresini aramadan roketin kapsülünü bulamıyorlar. Roketin kapsülünü bulmak kaç saat sürmüştür?
Nedensel: Harun, el yapımı roketini bilim fuarında sergileyecektir. Babası ve küçük kardeşi fuardan önceki gece, roketin parçalarını ayırıp paketlemesi için O'na yardım ederler. Harun'un babası, bir bahçe yapmak için evlerinin yanında, kırk dörde yetmiş altı metrelik küçük bir arsanın toprağını sürmüştür. Ertesi gün Harun'un küçük kardeşi, oyun oynarken, roketin kapsülünü bahçenin kenarına bir yere gömer. Harun ve kardeşi bir saat içinde bahçenin kenarları boyunca doğrusal olarak 40 metre uzunluğundaki bir yeri arayabilmektedirler. Bahçenin tüm çevresini aramadan roketin kapsülünü bulamıyorlar. Roketin kapsülünü bulmak kaç saat sürmüştür?
Olası Sonuç da içeren: Harun, el yapımı roketini bilim fuarında sergileyecektir. Babası ve küçük kardeşi fuardan önceki gece, roketin parçalarını ayırıp paketlemesi için O'na yardım ederler. Harun'un babası, bir bahçe yapmak için evlerinin yanında, kırk dörde yetmiş altı metrelik küçük bir arsanın toprağını sürmüştür. Ertesi gün Harun'un küçük kardeşi, oyun oynarken, bir roketin kapsülünü bahçenin kenarına bir yere gömer. Bilim fuarı beş saat içinde başlayacaktır. Harun ve kardeşi bir saat içinde bahçenin kenarları boyunca doğrusal olarak 40 metre uzunluğundaki bir yeri arayabiliyorlar. Bahçenin tüm çevresini aramadan roketin kapsülünü bulamıyorlar. Roketin kapsülünü bulmak kaç saat sürmüştür?

Nedensel öykü unsurlarının eklenmesi kimi problemlerde performansı düşürürken diğer bazı problemlerde artırmıştır. Ancak bu düşüş ve artışlar genellikle istatistikî anlamlılık düzeyine erişememiştir. Amerikalı ve Finli öğrenciler birbirlerine benzer bir davranış örüntüsü göstermişler ve genellikle nedensel unsurların eklendiği problemlerde görece daha yüksek performans göstermişlerdir. Diğer yandan Türk öğrenciler bütün problemlerde en yüksek performansı standart problemlerde göstermişlerdir. Ancak bunlardan sadece kurbağa probleminde performans farkları standart format lehine istatistikî olarak anlamlılık düzeyine ulaşmıştır. Özetle, nedensel öykü unsurlarının eklenmesi Amerikalı ve Finli öğrencilerin uzamsal içerikli problemlerdeki performansının artmasına neden olurken bu durum Türk öğrencilerde herhangi bir etki yapmamıştır.

Amerikalı ve Finli öğrencilerin benzer kültürlere sahip olmaları, okullarında Türk öğrencilerden farklı pratikler görmeleri, toplumsal olarak farklı değer yargılarına sahip olmaları bu çalışmada bu iki grup ülke öğrencilerden farklı sonuçlar alınmasına neden olmuş olabilir. Örneğin boşanma ile ilgili problemde hem Amerikalı hem de Finli öğrenciler nedensel unsurların eklendiği formatta daha yüksek performans göstermişlerdir. Diğer yandan Türk öğrencilerin daha çok standart problem formatlarında başarılı olmaları bulgusu Türkiye'de çoktan seçmeli test olarak yapılan merkezi sınavların etkisinden kaynaklanmış olabilir. Bu sınavlarda ve bu sınavlara hazırlık amacıyla yapılan çalışmalarda öğrencilere nesnellik kaygısıyla "steril" denebilecek, problemin çözümü için gerekli bilgiden fazla hiçbir bilginin bulunmadığı sorular sorulmaktadır. Böylece öğrenciler bir problemde fazladan bir bilgi verilmiş ise bunu gerekliymiş gibi algılayıp problemin çözümünü olduğundan daha zor olarak algılamış olabilirler. Sonuç olarak, nedensel öykü unsurlarının eklenmesinin farklı kültürlerde farklı etki yapabileceği bulgusundan hareketle sözel problemlerin çözümünde durumsal modellerin en az şema modelleri kadar öncelikli olabileceği sonucuna varılmıştır.