

# Effects of Linguistic Complexity and Math Difficulty on Word Problem Solving by English Learners

Otilia C. Barbu

Educational Psychology, The University of Arizona

PO Box 210069, Tucson AZ 85721 United States

Tel: 1-520-6152-984 E-mail: obarbu@email.arizona.edu

Carole R. Beal (Corresponding author)

Cognitive Science Program, The University of Arizona

PO Box 210025, Tucson AZ 85721 United States

Tel: 1-520-6269-860 E-mail: crbeal@email.arizona.edu

## **Abstract**

Prior research suggests that linguistic complexity may impede mathematics word problem solving by English Learners, but results have been inconsistent. The present study employed an experimental design to investigate the effects of linguistic complexity and mathematics difficulty on word problem solving by middle school English Learners. Results were consistent with predictions from Cognitive Load Theory: Performance was poorer for word problems written in more complex language compared to the same problems in easier text, and the weakest performance was observed for problems that were both linguistically and mathematically challenging. A Confirmatory Factor Analysis suggested a model including a latent factor, hypothesized to be working memory, provided a good fit to the data. Additionally, linguistic complexity had a significant influence on students' perceptions of the mathematical difficulty of the problems. The results are consistent with recent suggestions that English Learners' lower performance in math reflects the additional cognitive demands associated with text comprehension.

**Keywords:** Math problem solving, Cognitive Load Theory, English Learners, Middle school, Text characteristics

## 1. English Learners and mathematics achievement

The population of the United States has become more diverse in recent decades, and the challenges of educating students who are learning English have become increasingly apparent (Callahan, 2005; Capps, Fix, Murray, Ost, Passel & Herwanto, 2005; Villalba, Akos, Keeter & Ames, 2007). Much research has focused on the factors that contribute to the differences between English Learners and English Primary students in reading achievement (August & Shanahan, 2008; Batalova, Fix & Murray, 2007). There has also been growing awareness of the gap between English Learners and English Primary students in mathematics achievement. Recent research indicates that English Learners are less likely than English primary students to score well on math achievement tests, to succeed in algebra, and to achieve passing scores on mathematics exams that are required for high school graduation in many of the United States (National Center for Education Statistics, 2005; National Clearinghouse for English Language Acquisition, 2007). In the United States, resources for schools are now increasingly dependent on students' achievement test performance. Similar challenges are being faced in other nations with large proportions of students who are not yet proficient in the language of instruction (Hinkel, 2005). Thus, it is becoming critically important to understand more fully the nature of the challenges that are faced by students who are not yet proficient in English as they solve mathematics problems (Abedi, 2004; Capps et al., 2007; Rivera, Collum & Shafer Willner, 2006).

### 1.1 Theoretical perspectives

English Learners' mathematical performance is a multi-faceted issue, including challenges related to receiving instruction in a non-primary language and the relatively low rate of math teachers who have received training in working with English Learners (Coates, 2006). In addition, research must consider the possible confounding of students' language status with variations in socioeconomic status and other demographic factors. Multiple theoretical perspectives have been adopted in research on English Learners, including sociocultural-oriented analyses of the classroom context and a more cognitively-oriented theoretical framework.

#### 1.1.1 Classroom influences on math learning by English Learners

Some researchers have analyzed how the classroom context experienced by English Learners can impede or support mathematics learning. Cuevas (1984) noted that many English Learners are confronted with the need not only to master a new language but also to learn the distinct features of academic discourse. That is, learning English as it is used in an instructional context is likely to be considerably more demanding than acquiring basic conversational proficiency. In the specific case of mathematics learning, students must come to understand the mathematics "register," meaning the types of language used to convey mathematical concepts as well as the mathematics-specific meanings for words that may be familiar in other contexts (e.g., "table") (Enyedy, Rubel, Castellon, Mukhopadhyay, Esmonde & Secada, 2008; Moschkovitch, 2002, 2005).

Other researchers have noted that math instruction can sometimes rely on prior knowledge that may not always be shared by English Learners. For example, a student who is not familiar with American sports may not understand a problem about football scores or baseball

averages. These analyses make clear the need to ensure that teachers consider English Learners' experiences and take care to clarify forms of expression that may be confusing or ambiguous to a student who is learning English (Janzen, 2008). In addition, this theoretical perspective holds that it is not appropriate for math teachers to view English Learners as being "deficient" but rather to identify the strengths of their prior knowledge and experiences that can be integrated into the examples used in the classroom (Freeman & Crawford, 2008; Rodriguez, 2009; Secada, 1996).

### 1.1.2 English Learners' performance on mathematics achievement tests

Other researchers have focused more specifically on investigating why English Learners perform less well than English Primary students on math achievement tests. Recent studies have demonstrated that reading proficiency in English is a significant predictor of mathematics test performance by English Learners (Beal, Adams & Cohen, 2010; Guglielmi, 2008). In particular, the characteristics of math test items that are especially challenging for English Learners have been the subject of investigation (Echevarria, Powers & Short, 2006; Sireci & Khaliq, 2002; Solano-Flores, 2008). Martiniello (2008) identified a set of test items from a state math achievement test that were known to be challenging for English Learners, relative to English primary students with similar overall proficiency in math. She then worked with a small sample of English Learners as they attempted the items. The results pointed to several item characteristics that impeded the English Learners, including the use of low-frequency vocabulary terms (e.g., "spinner") as well as words that were ambiguous in meaning, such as "one" (which could be a pronoun or reference to the number) and "change" (a difference, or money received in a financial transaction). Other test items referred to concepts and examples that would not necessarily be familiar to English Learners from other cultures, such as coin-operated laundry equipment or a "spelling bee" competition. Although the linguistic complexity of math test items seems generally linked with poorer performance by English Learners, the reason for this effect is not yet entirely clear. Some findings suggest that the nature of the difficulty experienced by English Learners when solving math problems is actually more general or diffuse than an occasional issue with unfamiliar vocabulary. For example, Martiniello (2008) found that the overall length of the word problem, rather than the inclusion of unfamiliar vocabulary, was most consistently associated with poor problem solving for English Learners. Similarly, Wolf and Leon (2009) reported that the overall amount of academic vocabulary in word problem items was most predictive of item difficulty for English Learners. A meta-analysis of studies looking at the impact of various vocabulary-related accommodations in testing, including bilingual dictionaries, glossary definitions, dual language tests, and simplified English found surprisingly little evidence that such word-level accommodations consistently led to better performance for English Learners (Kieffer, Lesaux, Rivera & Francis, 2009).

### 1.1.3 Cognitive Load Theory

An alternative view was proposed by Campbell, Adams and Davis (2007) who noted that the challenge for English Learners in math problem solving goes beyond not understanding particular key words in the problems. These authors drew upon Cognitive Load Theory (CLT) to suggest that English Learners perform less well because they face additional overall cognitive demands during problem solving while working in a non-primary language.

Cognitive Load Theory holds that performance on complex cognitive tasks depends on whether the amount of information presented to the user equals or exceeds the availability of working memory; when working memory capacity is exceeded, the probability of errors will increase (Ashcraft, Donely, Halas & Vakali, 1992; Kalyuga, Ayres, Chandler & Sweller, 2003; Paas, Renkl & Sweller, 2003; Renkl & Atkinson, 2003; Royer, Tronsky, Chan, Jackson & Merchant, 1999; Swanson, Cooney & Brock, 1993; Sweller, 1994; Sweller & Chandler, 1994). Thus, if students must devote significant cognitive resources to text comprehension, fewer working memory resources will be available for mathematical problem solving, including identifying the appropriate math operation, forming the problem representation, performing computations and checking progress towards the solution. In the case of English Learners, the need to allocate cognitive resources to understanding a problem presented in a non primary language would reduce the resources available for math problem solving and in turn increase the probability of errors.

Consistent with the predictions of CLT, several studies of word problem solving by English Primary students have shown improved performance when simplified language is used, (Adams & Hitch, 1995; De Corte, Verschaffel & De Win, 1985; Fayol, Abdi & Gombert, 1987; Kintsch & Greeno, 1985). However, Davis-Dorsey, Ross & Morrison (1999) reported no benefits to students' problem solving when the language of the word problem was simplified.

In the case of English Learners, Abedi and Lord (2001) compared English Learners' performance on math test items in their original format and in versions revised with simplified language. Consistent with the predictions of CLT, English Learners performed better with the simplified language versions of the problems. However, the difference was actually quite slight in absolute terms (on average, the difference was equivalent to a fraction of an item on a test that included 10 problems) (Kieffer et al., 2009). Additionally, as noted earlier, the meta analysis of accommodations in testing showed little consistent benefit for English Learners associated with the use of simplified language (Kieffer et al., 2009). However, there have actually been relatively few studies investigating this issue. Shaftel, Belton-Kocher, Glasnapp and Poggio (2006) pointed to the need for additional investigation into the performance of English Learners on word problems that involve the same computational operations while varying in the complexity of the language.

#### 1.1.4 Predictions from CLT regarding linguistic complexity and mathematics difficulty

One prediction suggested by Cognitive Load Theory is that the English Learners should perform less well on hard-text problems relative to easy-text problems. This comparison was included in the present study to provide a needed replication of the results of Abedi and Lord (2001) as suggested by Shaftel et al. (2006).

A second prediction is that English Learners' problem solving performance should decline as the overall demands on working memory increase from multiple sources, including the demands of the mathematics operation required for the solution. More specifically, the poorest problem solving performance should be observed for word problems that are both linguistically complex and mathematically challenging. Although this prediction seems straightforward, there is surprisingly little prior research in which both the linguistic

complexity and the mathematical difficulty of the items have been varied independently. Studies of test items associated with disproportionately weak performance by English Learners have found that these items tend to involve both complex language and relatively difficult mathematics (e.g., probability), making it difficult to evaluate the contributions of each factor (Martiniello, 2008). In one study, researchers found that linguistically complex language had a negative effect on English Learners' performance only for relatively easy items, but not on the more difficult problems (Wolf & Leon, 2009). However, results were not reported separately for math and science items, making interpretation difficult. Because there has been only limited prior work in this area, both easy and difficult math word problems were included in the present study.

### 1.1.5 Goals of the study

The present study was conducted to address the need for a direct comparison of the performance of English Learners on math word problems that were constructed to vary systematically in both linguistic complexity and math difficulty. Math difficulty was crossed with linguistic complexity, meaning that each English Learner saw easy-math-easy-text problems, easy-math-hard-text problems, hard-math-easy-text problems, and hard-math-hard-text problems. Problem versions were counter balanced across students to ensure that differences in problem solving performance related to linguistic complexity or math difficulty were not due to the specific content of a particular item. The study design allowed us to evaluate two predictions based on CLT: First, performance should be poorer for word problems with more complex language relative to performance on problems with the same mathematical content and less complex language. Second, performance should decline on problems involving both complex language and more challenging math operations.

Multiple measures of problem solving were integrated into the task, with the objective of capturing indices of students' understanding of the problem (ability to identify the appropriate math operation) as distinct from their ability to solve the problem (perform the calculations correctly). In addition, students were asked to assess the demands of the problem in terms of its linguistic challenges (rating how easy it was to understand what the problem was asking) and mathematical challenges (rating how difficult it was to solve the problem). This procedure was suggested by Montague and Applegate (2000) who investigated the link between the perception of problem difficulty and problem solving performance by students with learning disabilities.

## 2. Study Method

### 2.1 Study participants

The initial study sample included 41 middle school (Grade 6 and 7) English Learners. Students were recruited from three middle schools in Tucson, Arizona. Students were eligible for participation in the study if they were identified by the district as an English Learner, were enrolled in the state-mandated English instruction period, and received written parental consent to participate. Forty students spoke Spanish as their primary language, and Chinese was the primary language for the other student.

### 2.2 Study materials

The study materials included paper booklets printed with eight math word problems: two

easy-math-easy-English, two easy-math-hard-English, two hard-math-easy-English, and two hard-math-hard-English problems. The easy math problems involved single-digit addition and simple multiplication. The difficult math problems involved multi-digit multiplication and division. The word problems focused on factual information about rare and endangered animals such as the Giant Panda, Great White Shark, California Condor, and others. One word problem was printed on each page, and included a small color picture of the animal. The picture did not include information that could have been used to solve the math problem. Each word problem was prepared in an easy-English and a hard-English version, with the two versions appearing in different booklets. For example, an easy math problem would appear in the easy-English version in one booklet, and the hard-English version in the second booklet. The text complexity was altered by changing the vocabulary and grammatical structure of the problem, while holding the overall word count constant. The difficulty of the word problem text was evaluated automatically by the REAP readability assessment software developed at Carnegie Mellon University < <http://reap.cs.cmu.edu> >. REAP considers both vocabulary frequency and grammatical complexity in its calculations of text difficulty, and has been established to be more effective than other readability metrics for relatively short texts (such as math word problems) (Heilman et al., 2008). The word problem texts were processed by REAP via its online portal, which returned an average readability of Grade 3 for the easy-English problems, and an average readability of Grade 8 for the hard-English problems. Examples are included in Appendix A.

### *2.3 Study procedure*

Each student was individually interviewed by one of two researchers, in a session that lasted about 30 minutes. Interviews were conducted in an empty classroom or office area near the student's classroom. Interviews were conducted in English.

When meeting with the student, the interviewer introduced the word problem booklet and explained that math problems could be difficult because the math was hard, because the English was difficult to understand, or for both reasons, and that the session goal was to review the math problems in the booklet and discuss the relative difficulty of the problems. The student was explicitly told that some problems were easy but others were quite challenging, that he or she was not necessarily expected to be able to solve them all without help, and that the interviewer would help the student find the solution if necessary. This procedure was adopted from prior research on comprehension monitoring, in which students are enlisted to assist the researcher by pointing out difficulties with understanding instructional materials (Markman, 1979). This approach is intended to ensure that students feel comfortable with describing any problems in comprehension.

On each problem, the interviewer read the word problem text out loud as the student followed along in the text. The interviewer then asked the student to indicate the math operation that the problem involved, e.g., addition, subtraction, multiplication, division, or something else? The student's initial answer was noted at the top of the page. If the student mentioned an incorrect operation, the interviewer then referred to the key area of the problem and explained the correct operation. A colored pen was used to mark the key area (e.g., a key word such as "double" or "average") in the problem text, which provided a record that a hint had been provided for that problem. If the student did not agree or said that he or she did not

understand, the interviewer continued by repeating and expanding the hint, again using a colored pen. For example, on a problem involving conversion from inches to feet the interviewer drew a picture to illustrate one foot and then marked off the foot into inches to show that the answer had to involve multiplication rather than division or some other operation. The interviewer then noted the correct operation at the top of the page.

Next, the interviewer asked the student to rate the difficulty of the words (meaning how easy it was to understand what the problem was asking) and the math (meaning how easy he or she thought it would be to solve the problem). The student used a highlighter to make his or her rating on a 5-part Likert-type rating scale printed below the problem text.

The student was then asked to solve the problem, working in pencil in an area below the problem text. If the student did not solve the problem correctly, the interviewer provided assistance, including helping the student set up the equation if necessary and pointing out computation errors or misplaced decimals. The interviewer referred to a worked solution in order to detect errors as the student worked, and used a colored pen to mark the error locations.

At the end of the interview, the student was asked to look back at the eight word problems in the booklet and indicate the one that he or she thought was most difficult.

#### *2.4 Scoring*

On each problem, the student received a score of 0 or 1 indicating whether the correct math operation had been identified. The number of hints provided by the interviewer for that problem was recorded (based on the notations made in colored pen) along with the number of unfamiliar vocabulary items that the student highlighted in the problem. Hints provided in the context of identifying the correct operation and setting up the right equation were considered errors of interpretation, whereas hints provided as the student solved the problem were to address errors of computation. A score of 0 or 1 was also recorded to indicate if the student solved the problem without assistance correctly or not.

Students also made two ratings of each problem: one for the perceived difficulty of the math operation, and the other for the perceived difficulty of the problem text. Ratings ranged from 1 (“very easy”) to 5 (“very hard”).

Each student completed eight problems, two each for easy-math-easy-English, easy-math-hard-English, hard-math-easy-English, and hard-math-hard-English. Scores and ratings were averaged across the two problems in each of these four categories for each student.

### **3. Results**

#### *3.1 Descriptive statistics for English Learners’ problem solving performance*

Mean scores for the math problem solving results (probability of identifying the correct operation for the problem, the number of hints per problem, and the number of computation errors per problem) are shown in Table 1. As may be seen in Table 1, students’ scores were generally best for the easy-text-easy-math problems, and poorest for the hard-text-hard-math problems.

Table 1. English Learners' math problem solving

	Math Difficulty	Correct Operation	Interviewer Hints	Computational Errors
Easy Text	Easy Math	0.82 (.27)	0.46 (.47)	1.82 (.74)
	Hard Math	0.57 (.40)	0.83 (.63)	1.87 (.87)
Hard Text	Easy Math	0.73 (.30)	0.72 (.55)	2.41 (.85)
	Hard Math	0.44 (.36)	1.16 (.81)	2.60 (1.01)

Mean scores per problem for probability of identifying the correct operation, number of hints provided by the interviewer, and number of computational errors. Standard deviations in parentheses.

We conducted two-way repeated measures analyses of variance with text complexity (easy, hard) and math difficulty (easy, hard) as the repeated factors for each of the three dependent variables. For correct identification of the required math operation on the problem, the ANOVA showed a significant main effect for math difficulty,  $F(1,37) = 21.79$ ,  $p < 0.001$ ,  $\eta^2 = 0.353$ . Students were more likely to identify the correct operation on problems that involved easy math. The main effect associated with text difficulty was also significant, but at a smaller value,  $F(1,37) = 4.83$ ,  $p = 0.034$ ,  $\eta^2 = 0.108$ . No significant interaction between factors was observed,  $F(1,37) = 0.067$ ,  $p = 0.796$ ,  $\eta^2 = 0.002$ .

For the number of hints provided by the interviewer as the student solved the problem, the ANOVA results indicated that there was a significant main effect associated with math difficulty,  $F(1,37) = 20.07$ ,  $p < 0.001$ ,  $\eta^2 = 0.334$ . The main effect associated with text difficulty was also significant  $F(1,37) = 14.62$ ,  $p < 0.001$ ,  $\eta^2 = 0.268$ . No significant interaction between the factors was observed,  $F(1,37) = 0.18$ ,  $p = 0.67$ ,  $\eta^2 = 0.005$ .

The results for the number of computational errors made by the students are reported next. When a computational error was detected during the problem solving process, the interviewer noted the error and helped the student to correct it before proceeding. Thus, students could make more than one error on a problem, but the errors were independent. The ANOVA indicated that there was a significant main effect for math difficulty,  $F(1,37) = 46.63$ ,  $p < 0.001$ ,  $\eta^2 = 0.538$ . Not surprisingly, students made more errors on problems that involved more challenging math operations. There was also a statistically significant, albeit smaller, main effect associated with the text difficulty of the problems,  $F(1,37) = 4.14$ ,  $p = 0.049$ ,  $\eta^2 = 0.094$ . In addition, there was a significant interaction between math difficulty and text difficulty,  $F(1,37) = 4.85$ ,  $p = 0.034$ ,  $\eta^2 = 0.108$ . When students worked on difficult math problems they made more computational errors when the problems were written in complex language than they did on the same problems written in relatively easy language. Recall that these errors included only computation mistakes, as errors regarding the identification of the correct math operation and setting up the appropriate equation had already been corrected

by the interviewer.

To summarize: English Learners generally performed best when they solved word problems written in relatively easy text that involved simple arithmetic operations. Their ability to identify the type of operation that the problem required was negatively affected by both the difficulty of the text and the difficulty of the math operation. With regard to computational errors made as the student actually solved the problem, the poorest performance was observed for problems with challenging text and difficult math. The overall pattern of performance was consistent with the predictions of CLT regarding demands on working memory.

### 3.2 Descriptive statistics for students' ratings of problem text and math difficulty

In addition to solving the problems, the students were asked to evaluate each problem in terms of its apparent difficulty with regard to understanding the text, and performing the computations. Students used a 5 point Likert-type rating score to record their ratings after solving each problem. Mean ratings are shown in Table 2.

Table 2. Students' ratings of math problem text and mathematical difficulty

	Math Difficulty	Math Rating	Text Rating
Easy Text	Easy Math	1.37 (.55)	1.82 (.74)
	Hard Math	2.10 (1.04)	1.87 (.87)
Hard Text	Easy Math	1.37 (.54)	2.41 (.85)
	Hard Math	2.54 (.89)	2.60 (1.01)

Mean student rating per problem for math difficulty and text difficulty. Ratings ranged from 1 ("very easy") to 5 ("very hard"). Standard deviations in parentheses.

Ratings for how easy it was to understand the problem were analyzed in a two way repeated measures ANOVA with text difficulty and math difficulty as the repeated factors. There was a main effect associated with text difficulty,  $F(1,37) = 41.51$ ,  $p < 0.001$ ,  $\eta^2 = 0.509$ , but no other factors were significant. Thus, students rated problems written in more challenging text as harder to understand, but math difficulty did not influence their ratings.

Ratings for how difficult the problem was to solve were analyzed in another ANOVA. There was a main effect of math difficulty,  $F(1,37) = 71.42$ ,  $p < 0.001$ ,  $\eta^2 = 0.641$ . Not surprisingly, when the math problem involved relatively challenging mathematics (e.g., multi digit division), students rated it as more difficult to solve than problems involving basic addition. Interestingly, the analysis also revealed a main effect associated with text difficulty,  $F(1,37) = 6.64$ ,  $p = 0.014$ ,  $\eta^2 = 0.142$ . This indicates that when the problem was written in linguistically complex text, students perceived it to be mathematically more

challenging than when the same problem was written in simple text. There was also a significant interaction,  $F(1,37) = 6.01$ ,  $p = 0.019$ ,  $\eta^2 = 0.131$ . Not surprisingly, students rated problems with both challenging math and complex language as most difficult.

To summarize: The main effects in these analysis indicated that English Learners could clearly distinguish the sources of difficulty in a problem (text vs. math difficulty). The results also revealed that students' impressions of the mathematical difficulty of the problem were influenced by the complexity of the text: Problems written in harder text appeared to be more difficult to solve than the same problems presented in easier text. Text difficulty also influenced students' choice of the most difficult problem in the booklet: 53% of the sample chose a hard-math-hard-English problem, 11% chose an easy-math-hard-English problem, and 36% chose a hard-math-easy-English problem. No student selected an easy-math-easy-English problem as the most difficult.

### 3.3 Correlations among dependent variables

Correlations between the dependent variables are shown in Table 3. First, we note that students' ratings of the math difficulty of the problems are strongly correlated with the number of computational errors made on the problems ( $r = 0.514$ ,  $p < 0.01$ ). Math difficulty and text difficulty ratings were correlated, ( $r = 0.481$ ,  $p < 0.01$ ), which is consistent with the finding that linguistic complexity affects students' perception of the mathematical difficulty of the problem. The correct operation identification variable correlates negatively at a medium level with the math difficulty rating variable ( $r = -0.345$ ,  $p < 0.01$ ), but only at a small negative level with the text difficulty rating ( $r = -0.235$ ,  $p < 0.01$ ). Because the math difficulty rating is influenced by the text difficulty, we cannot conclude for certain whether math or text difficulty has the most influence on students' ability to identify the correct math operation.

Table 3. Correlations between dependent variables

	Operation	Hints	Errors	Text Rating	Math Rating
Operation	1	0.433**	-0.181*	-0.235**	-0.345**
Hints		1	-0.644**	0.444**	0.534**
Errors			1	0.164	0.514**
Text Rating				1	0.481**
Math Rating					1

Correlation matrix between observed variables for correct operation identification, number of hints, errors, text rating and math rating. Note: (\*\*)  $p < 0.01$  two-tailed test, (\*)  $p < 0.05$  two-tailed test.

Finally, we see that there are medium and large level correlation effects between the number

of hints required by the students and the other four variables investigated. More hints were needed if the student could not easily identify the requirements of the problem. The correlations were positive for math and text difficulty ratings and the number of computational errors made on the problems.

### 3.4 Confirmatory factor analysis (CFA)

The results are generally consistent with the predictions of CLT, but no direct measure of working memory during problem solving performance was included in the study. To determine if a model based on working memory as a latent variable would be a reasonable interpretation of the results, a confirmatory factor analysis was conducted using AMOS package included with SPSS (version 16.0). The principal component analysis performed on the variables indicated in Table 3 revealed the existence of only one significant eigenvalue 2.64 responsible for 53% of total variance; thus, the Kaiser-Guttman rule would suggest a unidimensional latent structure. The eigenvalues corresponding to the other four components were smaller than 1 and they were neglected (Kaiser, 1960). Model 1 has two levels and contains the latent variable hypothesized to be “Working Memory.” The latent variable is affected by text and math difficulties and is considered to control the observed variables “Math difficulty rating” and “Number of computational errors.” In contrast, Model 2 considers that all 4 observed variables are directly controlled by the text and math difficulty levels of the problems.

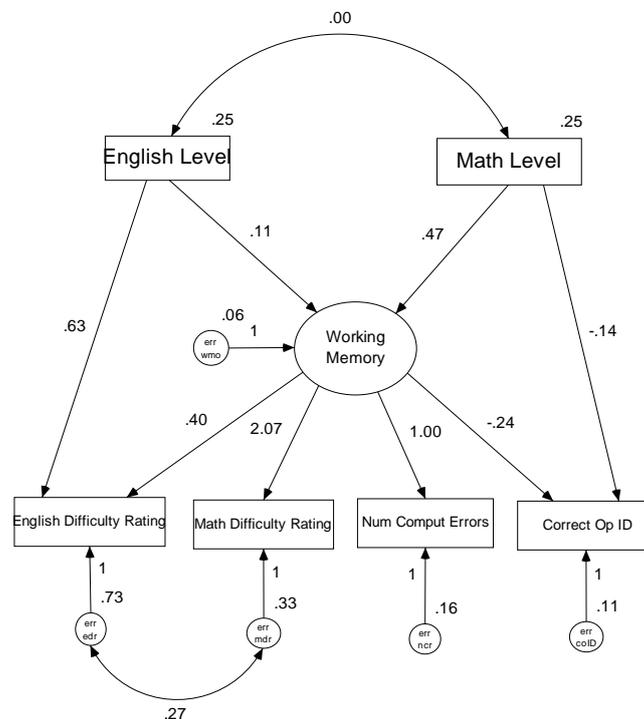


Figure 1. Model 1 with Working Memory as a latent variable. CMIN/df = 1.74, GFI = 0.98, TLI = 0.96, CFI = 0.99, RMSEA = 0.067.

In Figures 1 and 2 we show the two models that have been compared: Model 1 has two levels and contains the latent variable called “Working Memory”, which is affected by

English/Math level difficulties and is considered to control the observed variables “Math difficulty rating” and “Number of computational errors”. In contrast, Model 2 considers that all 4 observed variables are directly controlled by the English/Math level difficulties.

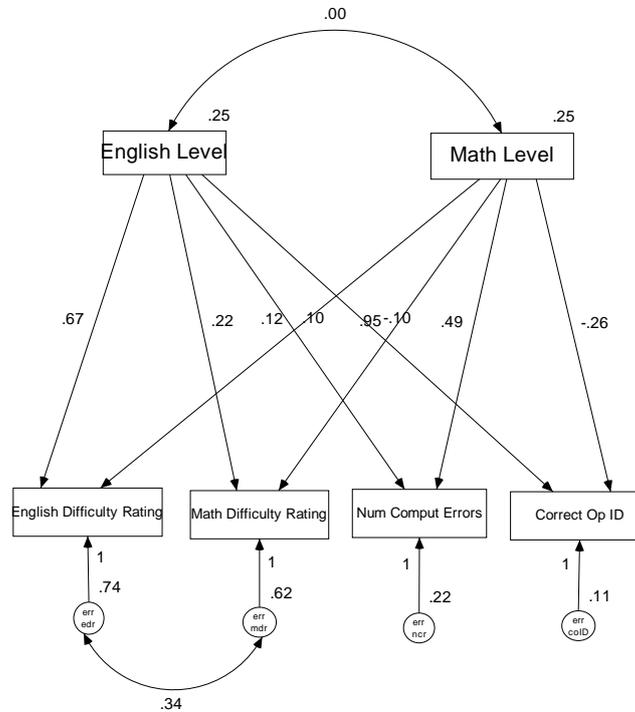


Figure 2. Model 2 in which English and Math difficulty levels have direct influence on all the observed variables CMIN/df = 5.89, GFI = 0.95, TLI = 0.73, CFI = 0.88, RMSEA = 0.17.

Because many fit statistics are possible for the model evaluation, we based our decision on recommendations made by Hu and Bentler (1999). For the chi-square statistics (CMIN) we report the chi-square difference between the observed data and the specified model. This value is influenced by the number of degree of freedom (df), so we also report the value of chi-square divided by the number of degrees of freedom (CMIN/df) with a value less than 2 indicating a good fit. Next we report the Goodness of Fit Index (GFI), the Tucker–Lewis coefficient (TLI), and the Bentler Comparative Fit Index (CFI). Following Hu and Bentler (1999) a value larger than 0.9 for these 3 fit indexes is considered a good fit and a value larger than 0.95 an excellent fit. We also report the root mean square error of approximation (RMSEA) for which a value less than 0.05 is considered a good fit and a value less than 0.8 is considered a reasonable fit (Browne & Cudeck, 1992). A value of RMSEA larger than 0.1 is an indicator of a poor fit. Finally, the Akaike’s Information Criterion (AIC), which is a function of both model complexity and goodness of fit will be used for a direct comparison of the models (the lower the score the better the model). The values for these indexes are shown in Table 4.

Table 4. Comparative fit indexes for two proposed models

	CMIN	Df	CMIN/df	GFI	TLI	CFI	RMSEA	AIC
Model 1	10.42	6	1.74	0.98	0.96	0.99	0.067	54.43
Model 2	35.32	6	5.89	0.95	0.73	0.88	0.17	79

Model 1 represents working memory as a latent factor; Model 2 assumes text and math difficulty influence observed variables directly.

Model 1, which considers the effects of both text and math difficulty on the hypothesized working memory latent variable, meets the criteria for a very good fit of the experimental data. The value of  $CMIN/df = 1.74$  is less than 2 and corresponds to a value of  $p = 0.11$  and so the null hypothesis (which verifies the accuracy of the suggested model) can not be rejected. The 3 fit indexes considered (GFI, TLI and CFI) are all higher than 0.95 and the value of  $RMSEA = 0.067$  indicates a reasonable fit. Finally, the AIC value (54.43) indicates that overall this model performs better than the Model 2. In contrast, the second model, which does not consider the hypothesized effects of working memory overload, does not meet all the criteria for a good fit of the experimental data. The value of  $CMIN/df = 5.89$ ,  $p < 0.001$  leads to the rejection of the null hypothesis, while the values of two of the fit indexes (TLI and CFI) are below the cut-off value of 0.9. Similarly, the value of  $RMSEA = 0.17$  indicates a poor fit of the data. Thus, the results support the interpretation that students' problem solving performance was related to a latent factor, hypothesized to be working memory, that was influenced by the primary manipulations in the experiment (linguistic complexity and math difficulty of the word problems).

## 4. Conclusion

### 4.1 Summary of results in relation to predictions of Cognitive Load Theory

The study was designed to provide an experimental comparison of how linguistic complexity influenced English Learners' problem solving on items that involved easy or challenging mathematics. Hypotheses suggested by Cognitive Load Theory were that students should perform less well on word problems written with complex language relative to the same problems written in easier language, and that performance should be weakest on problems involving both complex language and difficult mathematics. Middle school English Learners were presented with word problems constructed to vary in math operation (very easy or relatively challenging) and in the readability of the text (Grade 3 or Grade 8 average readability).

Consistent with the first hypothesis, the results showed that the English Learners in the study performed well when the math problems involved easy operations such as simple addition and multiplication; they solved over 90% of these items correctly. In contrast, their performance declined when they were asked to solve word problems that involved more challenging mathematics operations (multi-digit multiplication and division). Students

solved only about half of these problems correctly. Prior research that had found only modest or inconsistent effects of linguistic complexity but the language factor had been confounded with the difficulty of the mathematics operation. The experimental design of the present study allowed us to document clearly that when the text is relatively difficult to understand due to higher grammatical complexity and more advanced vocabulary, English Learners were less successful in problem solving than when the text is written in simplified English, even though the actual math operation was the same. The negative impact of difficult text was apparent in multiple measures of problem solving, including students' ability to identify the appropriate math operation for the problem as well as their ability to perform the correct computations. Students required more help and made more errors on problems written in challenging English than similar problems expressed in easy English text, even when the actual difficulty of the mathematics was held constant.

Results also provided support for the second hypothesis: Students showed the poorest performance on problems that involved both complex text and relatively difficult mathematics. Although this result is not particularly surprising, the additive impact of the language and mathematics factors had not previously been demonstrated. In addition, the study design allowed us to investigate the impact of linguistic complexity on students' perceptions of the mathematics of the problem, meaning how difficult it would be to solve. Students were asked to provide independent ratings of the linguistic complexity and the mathematical difficulty of each problem. Their ratings varied appropriately, suggesting that students were sensitive to the variations in the problem characteristics.

The findings are generally consistent with the analysis of Campbell et al. (2007) who pointed to the additional cognitive demands faced by English Learners when solving math problems in a non-primary language. However, the conclusions are limited because there was no direct assessment of the students' actual working memory capacity or the impact of the problem characteristics on working memory. The results of the confirmatory factor analysis pointed to the influence of one latent factor on performance, but we cannot be sure that the latent factor actually represents students' working memory. Additional research will be needed to investigate more directly how the cognitive demands of understanding a math problem interact with the processes required to solve it.

#### *4.2 Broader implications of the findings*

The present study also revealed that linguistic complexity interacts with math difficulty to influence English Learners' perception of the problems. More specifically, math problems written in complex language were rated as mathematically more difficult to solve than the same problems written in simpler language. These results raise the possibility that when English Learners are asked to solve math problems that are difficult to understand because of the text, they may become discouraged because the problem appears harder than it actually is. This result may help to inform related research on English Learners' mathematics self-concept. Some prior work indicates that the relation between prior math performance and math self-concept is stronger for Hispanic students than White students (Stevens, Olivarez, Lan & Tallent-Runnels, 2004). Although it was not reported how many of the Hispanic

students in the Stevens et al. study were English Learners, the present findings suggest that struggling to solve math problems that appear to be difficult because of the text characteristics may influence English Learners' belief that they can do well in math.

The present findings raise an additional concern related to the validity of tests administered to English Learners. If the language complexity of the problem interferes with English Learners' ability to make an accurate assessment of its mathematical difficulty, their performance on achievement tests may be negatively affected. Students might even avoid attempting problems that they could have solved. Also, if the language of the problem affects students' ability to identify the operation needed to solve it and increases the probability that they will make computation errors then the item is assessing more than students' mathematical proficiency. Thus, the present results add to other research suggesting that language complexity may undermine the validity of test items designed to assess mathematical problem solving when used with English Learners (Abedi, 2004; Kieffer et al., 2009; Solano-Flores, 2008). Additional research is needed to investigate these issues.

### **Acknowledgments**

We would like to thank the teachers, principals and district administrators of the Amphitheater School District in Tucson AZ for their support of the research. We would also like to thank Dr. Jamie Callan of the Language Technologies Institute at Carnegie Mellon University for allowing us to use the REAP software to evaluate the readability of the word problems used in the study. The research was supported by a grant from the National Science Foundation (DRL 0903441). The views expressed here are not necessarily those of the sponsoring agency.

### **References**

- Abedi, J. (2004). The No Child Left behind Act and English Language Learners: Assessment and accountability issues. *Educational Researcher, 1*, 4-14.
- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*, 219-234.
- Adams, J. W., & Hitch, G. J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology, 67*, 21-38.
- Ashcraft, M. H., Donley, R. D., Halas, M. A., & Vakali, M. (1992). Working memory, automaticity, and problem difficulty. In J. I. D. Campbell (Eds.), *The nature and origins of mathematical skills* (vol. Advances in Psychology, 91, pp. 301-329). Amsterdam: Elsevier.
- August, D., & Shanahan, T. (2008). *Developing reading and writing in second language learners*. New York NY: Routledge.
- Ayers, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*, 389-400.
- Batalova, J., Fix, M., & Murray, J. (2007). *Measures of change: The demography and literacy of adolescent English language learners*. New York: National Center on Immigrant

Integration Policy.

Beal, C. R., Adams, N., & Cohen, P. R. (2010). Reading proficiency and mathematics problem solving by high school English Language Learners. *Urban Education, 44*, 58-74.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21*, 230–258.

Callahan, R. (2005). Tracking and high school English Learners: Limiting opportunity to learn. *American Educational Research Journal, 42*, 305-328.

Campbell, A. E., Adams, V. M., & Davis, G. E. (2007). Cognitive demands and second language learners: A framework for analyzing mathematics instructional contexts. *Mathematical Thinking and Learning, 9*, 3-30.

Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Herwanto, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington DC: Urban Institute.

Coates, G. D. (2006). *Closing the achievement gap in mathematics for EL students*. Presentation at Summit V, Office of English Language Acquisition, Washington DC.

Cuevas, G. J. (1984). Mathematics learning in English as a second language. *Journal for Research in Mathematics Education, 15*, 134-144.

Davis-Dorsey, J., Ross, S. M., & Morrison, G. R. (1999). The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology, 83*, 61-68.

De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology, 77*, 460-470.

Echevarria, J., Powers, K., & Short, D. (2006). School reform and standards-based education: A model for English Language Learners. *Journal of Educational Research, 4*, 195-210.

Enyedy, N., Rubel, L., Castellon, V., Mukhopadhyay, S., Esmonde, I. & Secada, W. (2008). Revoicing in a multilingual classroom. *Mathematical Thinking and Learning, 10*, 134-162.

Fayol, M., Abdi, H., & Gombert, J. E. (1987). Arithmetic problems formulation and working memory load. *Cognition and Instruction, 4*, 187–202.

Freeman, B., & Crawford, L. (2008). Creating a middle school mathematics curriculum for English-Language Learners. *Remedial and Special Education, 29*, 9-19.

Guglielmi, R. S. (2008). Native language proficiency, English literacy, academic achievement, and occupational attainment in Limited-English-Proficient students: A latent growth modeling perspective. *Journal of Educational Psychology, 100*, 322-342.

Heilman, M., Collins-Thompson, K., & Eskenazi, M. (2008). *An analysis of statistical models and features for reading difficulty prediction*. The 3rd Workshop on Innovative Use of

NLP for Building Educational Applications. Association for Computational Linguistics.

Hinkel, E. (Ed.). (2005). *Handbook of research in second language teaching and learning*. Mahwah NJ: Erlbaum.

Hu, L., & Bentler, P. M. (1999). Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

Janzen, J. (2008). Teaching English Language Learners in the content areas. *Review of Educational Research*, 78, 1010-1038.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38, 23-31.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English Language Learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, 79, 1168-1201.

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

Markman, E. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, 50, 643-655.

Martiniello, M. (2008). Language and the performance of English-Language Learners in math word problems. *Harvard Educational Review*, 78, 333-368.

Montague, M., & Applegate, B. (2000). Middle school students' perceptions, persistence and performance in mathematical problem solving. *Learning Disability Quarterly*, 23, 215-227.

Moschkovitch, J. (2005). Using two languages when learning mathematics. *Educational Studies in Mathematics*, 64, 121-144.

Moschkovitch, J. (2002). A situated and sociocultural perspective on bilingual mathematics learners. *Mathematical Thinking and Learning*, 4, 189-212.

National Center for Education Statistics. (2005). *National Assessment of Educational Progress*. Washington DC: U.S. Department of Education, Institute of Education Sciences.

National Clearinghouse for English Language Acquisition. (2007). [Online] Available: [http://www.ncela.gwu.edu/stats/2\\_nation.htm](http://www.ncela.gwu.edu/stats/2_nation.htm) Sept. 28, 2007.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist*, 38, 1-4.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist*, 38, 15-22.

- Rivera, C., Collum, E., & Shafer Willner, L. (2006). *State assessment policy and practice for English Learners: A national perspective*. Mahwah NJ: Lawrence Erlbaum.
- Rodriguez, D. (2009). Meeting the needs of English Language Learners with disabilities in urban settings. *Urban Education, 44*, 452-464.
- Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Merchant, H. (1999). Math fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology, 24*, 181-266.
- Secada, W. G. (1996). Urban students acquiring English and learning mathematics in the context of reform. *Urban Education, 30*, 422-448.
- Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English Learners and students with disabilities. *Educational Assessment, 11*, 105-126.
- Sireci, S. G., & Khaliq, S. N. (2002, April). *An analysis of the psychometric properties of dual language test forms*. Paper presented at the annual meeting of the National Council on Measurement in Education. New Orleans LA.
- Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English Language Learners. *Educational Researcher, 37*, 189-199.
- Stevens, T., Olivarez, A., Lan, W. Y., & Tallent-Runnels, M. K. (2004). Role of mathematics self-efficacy and motivation in mathematics performance across ethnicity. *Journal of Educational Research, 97*, 208-221.
- Swanson, H. L., Cooney, J. B., & Brock, S. (1993). The influence of working memory and classification ability on children's word problem solution. *Journal of Experimental Child Psychology, 55*, 374-395.
- Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 295-312.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185-233.
- Villalba, J. A., Akos, P., Keeter, K., & Ames, A. (2007). Promoting Latino student achievement and development through the ASCA National Model. *ASCA Professional School Counseling, 10*, 464-474.
- Walczyk, J. J., & Griffith-Ross, D. A. (2006). Time restriction and the linkage between subcomponent efficiency and algebraic inequality success. *Journal of Educational Psychology, 98*, 617-627.
- Wolf, M. K., & Leon, S. (2009). An investigation of the language demands in content assessments for English Language Learners. *Educational Assessment, 14*, 139-159.

---

**Appendix**

1. Example of Easy math, easy English problem (Word count: 44; Readability: Grade 4)

The Snow Leopard has a very long tail. It wraps its tail around its face to keep warm. Its tail and its body are the same length. If its tail is 2 feet long, how many feet long is the Snow Leopard in all?

2. Example of Easy math, hard English problem (Word count: 42, Readability: Grade 9)

The Snow Leopard has an unusually long tail, which it often uses to shield its face for warmth. Its tail and body are equivalent in length. If its tail is 2 feet long, what is the total length of the Snow Leopard?

3. Example of Hard math, easy English problem (Word count: 40, Readability: Grade 3)

There are no Wild Horses living in the wild any more. But there are 2,100 Wild Horses living in zoos. There are 175 zoos around the world that have Wild Horses. How many horses does each zoo have on average?

4. Example of Hard math, hard English problem (Word count: 41, Readability: Grade 10)

The Wild Horse species has been extinct in the wild for several decades. However, there are 2,100 Wild Horses in captivity in zoos. If 175 zoos have Wild Horses in their collections, what is the average number of horses per zoo?

**Copyright Disclaimer**

Copyright reserved by the author(s).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).