Problem Solving by English Learners and English Primary Students in an Algebra-readiness ITS

Federico Cirett and Carole R. Beal

University of Arizona
Computer Science Dept., Cognitive Science Program
Tucson AZ 85721 USA
fcirett@email.arizona.edu, crbeal@email.arizona.edu

Abstract

Problem solving was compared for English Learners and English Primary students who had used AnimalWatch, an algebra-readiness ITS. Data records for word problems solved by students in both language groups were located and compared. Results indicated that English Learners were less likely to answer correctly, had more incorrect answer attempts, took longer on each problem, and were more likely to use multimedia help features in the ITS. Differences between English Learners and English Primary students were most apparent on word problems with challenging English text.

Introduction

A major challenge in research on intelligent tutoring systems (ITS) is how to evaluate the general impact of such systems on students when the instruction is, by definition, individualized. More specifically, when an ITS is integrated into the classroom, all students may work with the ITS but their individual instructional experience is likely to be unique. The diversity of instruction results because the ITS problem selector responds to student behaviors such as incorrect answers, requests to use multimedia help resources, and "gaming" activity. Thus, students who make many errors or attempt to "game" the system while solving problems will make slower progress through the ITS curriculum than those with better problem solving.

The challenges of ITS evaluation are especially apparent when variation in individual student performance is related to group characteristics. We might compare the outcomes of two groups that used an ITS and find that one group out-scored the other. However, if students in the two groups did not receive the same instruction, such comparisons may be misleading and a deeper investigation may be warranted (Beal & Cohen, 2005). For example, prior work has indicated that students who are not proficient speakers of English tend to perform less well in math problem solving than students whose primary language is English. If English Learners made more errors on AnimalWatch math word problems due to difficulties with understanding the English-based materials, then they would move more slowly through the ITS curriculum. As a result, they would be less likely to reach all the material, and might be less likely to show improvement on a post test. In effect, if the instruction provided by the ITS was less accessible to some students than others, it would be important to evaluate the impact separately for the two groups, as well as to identify factors that might affect some students more than others.

In the situation reported here, we had conducted a study with high school students who used AnimalWatch, an algebra-readiness ITS focusing on word problem solving, as part of their algebra class instruction (Beal, Adams & Cohen, 2010). The school district subsequently asked us to compare the performance of the students who were English Learners (ELs) and those whose primary language was English (EPs), and to look at the possible effects related to reading proficiency on the math problem solving of English Learners.

Math achievement by English Learners. Prior research indicates that English Learners in the United States generally perform less well than English Primary students in mathematics achievement (for reviews see August & Shanahan, 2006; Kieffer, Lesaux, Rivera & Francis, 2009). English Learners score lower on state achievement tests; are less likely to pass algebra; are less likely to obtain passing scores on math exams required for high school graduation; and are at greater risk for not completing high school. The challenges faced by English Learners are exacerbated by the fact that only about 15% of math teachers receive training in how to help ELs (Coates, 2006). In addition, many states in the United States have policies that emphasize English-only instruction, meaning that students may not receive instruction in their primary language or may do so for only a brief transitional period.

Theoretical framework. Although the relatively low math performance of ELs has been identified as an

Copyright © 2010, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

important research topic, the research base is relatively limited. In particular, there has been relatively little research to assess whether English Learners' ability to read the text of a math problem might influence their math math problem solving. One reason that reading proficiency might influence math problem solving is suggested by the cognitive resources theoretical framework outlined by Royer and colleagues (Royer et al., 1999). When students must devote cognitive resources to relatively low level problem solving operations, they will have fewer resources available to devote to higher order processes such as forming the appropriate problem representation, selecting a good problem solving strategy, and monitoring progress to the solution. More specifically, when students must work under time pressure, or when they have not mastered basic math facts, their performance falters. Conversely, additional practice designed to build proficiency with basic math facts is predictive of improved problem solving performance (Royer et al., 1999).

To date, the cognitive resources framework has not directly addressed the potential impact of text comprehension and reading proficiency on math problem solving. However, it seems reasonable to suggest that if a student must struggle to understand the words of the problem text, he or she would have fewer cognitive resources available to allocate towards the problem solving operations, and that increased errors would be likely. To evaluate this possibility, we conducted a post-hoc analysis of the ITS problem solving data, looking for cases in which English Learners and English Primary students had solved the same math word problems, and looking specifically for cases in which we could evaluate the impact of the English text on students' problem solving.

To answer the school's query, the student problem solving data had to be supplemented with other data provided by the district, specifically, information about the students' language status. In addition, the new focus on English Learners meant that we had to extend the information about each of the problems included the ITS. More specifically, AnimalWatch focuses on word problem solving, generally recognized to be an important component of mathematics proficiency (Koedinger & Nathan, 2004). However, because the AnimalWatch problem representations did not originally include readability metrics for the word problems, this information had to be added to support the analyses. Finally, we needed to locate word problems that had been solved by both English Learners and English Primary students. The analyses thus involved post-hoc alignment of multiple data sets and enrichment with new information, following by mining of the data to locate comparable cases.

Method

Participants

Participants included Grade 9 EL and EP students who had worked with the AnimalWatch ITS as part of their algebra math class instruction (N = 442, 209 ELs and 233 EPs). Teachers wanted their students to work with the ITS for review of pre-algebra material. The students attended schools in downtown Los Angeles CA. Student mathematics proficiency in the schools was generally low; 86% of the students scored below the "Basic" level of proficiency on the end-of-year California math achievement test.

AnimalWatch ITS

Student problem solving data were automatically collected as the students worked with the AnimalWatch tutoring system. AnimalWatch presents the student with word problems involving authentic science content about endangered species and environmental science topics (www.animalwatch.org). Word problems are organized into thematic narratives to provide students with a sense of progress as they solve a series of problems (e.g., "plan a trip to view Giant Pandas in the Wolong Reserve in China").

Within each thematic narrative, the problem selector algorithm searches for candidate problems that involve the current math topic (e.g., single digit multiplication) and selects a problem that is estimated to be within the student's ability to solve. If the student successfully solves the problem, the problem selector will attempt to choose more difficult problems or move on to a more challenging math topic (e.g., multi digit multiplication). The particular problems shown to a student are selected in response to the student's behavior (e.g., getting the answer correct, making one or more errors, etc.) and to the student's choice of theme (e.g., Pandas, Right Whales, Sharks, etc.). Thus, the particular sequence of problems shown to a student is highly individualized.

On each AnimalWatch word problem, students can enter up to three incorrect answers; a fourth incorrect answer will elicit the correct answer along with a message indicating that the student has failed to solve the problem. Students can also choose to move on to the next problem at any time (although problem "skipping" is generally very infrequent). While working on a word problem, students can also choose to view multimedia help resources at any time, including interactive examples (e.g., find the least common multiple of two denominators) and video lessons.

Data sources

ITS problem solving records. Records of students' problem solving while working with the ITS were located in the AnimalWatch database. These records consisted of time-stamped flat files representing the sequence of actions on each problem (e.g., problem is presented, student enters answer after some latency, answer is evaluated as correct or incorrect, student clicks "help" icon, answer is entered, answer is evaluated as correct, student clicks "next problem" icon, etc.).

The primary metrics of problem solving for the present analyses were the total time on each problem (number of seconds), the number of incorrect answer attempts on the problem, if the correct answer was ever entered, and whether the multimedia hints were accessed during the problem. The data set was also limited to students who had solved at least 10 problems in the ITS.

English Learners. The school district provided a data file that listed the students who had scores on the California English Language Development Test (CELDT, a state-specific instrument used to assess proficiency with English). This data source allowed us to identify the ELs in the sample of students who had worked with the ITS. No comparable data were available for the English Primary students (who did not take the CELDT test).

Word problem readability. The problem solving data for each student included a record of the specific problems shown to that student (i.e., the sequence of problem IDs). The original problem metrics included the math topic (e.g., addition) and difficulty (easy, average, challenging) of the problem. To add information about the problem text, every problem viewed by a student in the sample was located in the ITS database. The text of these problems was then processed to obtain a grade-level readability metric from the REAP software developed at the Language Technologies Institute at Carnegie Mellon University (reap.cs.cmu.edu). The REAP metric was selected because it includes assessments of grammatical complexity as well as vocabulary frequency, and has been shown to be relatively successful with short texts (such as math word problems). A script was written to submit the word problem text to the REAP online portal and store the returned readability metric with the problem.

Selection of word problems for analysis. The next step was to locate problems in the data set that appeared in both English Learners' and English Primary students' problem solving sequences, and to reformat the data set so that we could compare performance of the two student groups on specific word problems. We decided on the constraint that a problem had to have been presented to at least 20 English Learners and at least 20 English Primary students, to support the planned statistical analyses. The final data set included 5,188 problem solving records for 87 word problems.

Results

English Learners vs. English Primary students

We first investigated the performance of the two groups of students on AnimalWatch word problems that had been viewed by both ELs and EPs. Mean scores and standard deviations for the primary measures (time per problem; number of incorrect answers; whether the problem was ever solved correctly; and whether hints were activated on the problem) are shown in Table 1 for English Learners and English Primary students.

T-tests were conducted with student language group (EL, EP) as the grouping factor. The results indicated that ELs took significantly longer per problem than EPs ($t(5186) = 3.54$, $p < .001$); made more incorrect answer attempts on each problem ($t(5186) = 7.78$, $p < .001$); were less likely to enter the correct answer on a problem ($t(5186) = 8.71$, $p < .001$); and were more likely to activate the multimedia help features on a problem ($t(5186) = 5.93$, $p < .01$).

	Time (secs)	Incorrect Attempts	Correct Answer	Help Used
ELs	98.67 (114.17)	1.76 (1.28)	0.75 (.42)	0.11 (.31)
EPs	88.24 (97.87)	1.50 (1.04)	0.85 (.35)	0.06 (.24)

Table 1. Mean time, number of incorrect answers, correct solutions, and hint activation per problem for English Learners and English Primary students. Standard deviations are shown in parentheses.

The results of these initial analyses were not particularly surprising, given that prior studies of mathematics achievement in classroom contexts had found that ELs tend to show worse performance than EPs (Kieffer et al., 2009). However, the analyses did confirm that this difference was observed for AnimalWatch problems that both ELs and EPs had completed.

Easy text vs. hard text word problems

The next step was to learn if the readability of the word problem text influenced students' problem solving. The readability metrics added to the word problems were used to group the problems into two categories: Problems with text that required Grade 7 or higher reading skills, and problems with text that required reading skills under Grade 7. Grade 7 level text was chosen because this split point produced roughly equivalent numbers of word problems in the set of available problems.

Table 2 shows the mean time (in seconds) per problem for the easy-text and hard-text problems completed by English Learners and English Primary students.

	Easy text	Hard text
Eng Learners	93.78 (102.59)	109.74 (86.22)
English Primary	84.81 (90.47)	96.19 (112.84)

Table 2. Mean seconds per problem for easy and hard text problems. Standard deviations are shown in parentheses.

An analysis of variance with text (easy, hard) and student language group (EL, EP) as grouping factors indicated that the easy text problems were completed in significantly less time than the hard text problems, $F(1,5184) = 17.937$, $p < .001$. Also, the ELs took longer on average on the problems than the EPs, $F(1,5184) =$

12.425, $p < .001$. However, the interaction of text and language group was not significant; that is, both ELs and EPs took longer when the problem text was relatively hard.

The number of incorrect answers made on easy and hard text problems is shown in Table 3. An analysis of variance indicated that significantly more errors were made on hard-text problems than easy-text problems, $F(1,5184) = 14.61$ $p < .001$, and by ELs than by EPs, $F(1,5184) = 60.42$, $p < .001$. The interaction of language group and text difficulty was not significant.

	Easy text	Hard text
Eng Learners	1.71 (1.27)	1.81 (1.27)
English Primary	1.47 (1.05)	1.58 (1.01)

Table 3. Mean number of incorrect answer attempts per problem for easy and hard text problems. Standard deviations are shown in parentheses.

As may be seen in Table 4 below, problems solved by ELs were more likely to have had multimedia help activated than problems solved by EPs, $F(1,5184) = 35.21$, $p < .0001$. However, there were no effects related to text difficulty.

	Easy text	Hard text
Eng Learners	0.10 (.31)	0.11 (.31)
English Primary	0.06 (.25)	0.05 (.22)

Table 4. Mean proportion of problems on which help features were activated. Standard deviations are shown in parentheses.

Table 5 shows the mean proportion of problems correctly solved (including cases preceded by incorrect answer attempts). Hard-text problems were less likely to be solved correctly, $F(1,5184) = 23.07$, $p < .001$. Correct solutions were less frequent on problems solved by ELs compared to problems solved by EPs, $F(1,5184) = 75.76$, $p < .001$. Although the lowest score was for ELs on hard-text problems, the interaction was not significant.

	Easy text	Hard text
Eng Learners	0.77 (.41)	0.70 (.45)
English Primary	0.86 (.34)	0.82 (.38)

Table 5. Mean proportion of problems solved correctly for easy and hard text problems. Standard deviations are shown in parentheses.

These results documented that students generally did not perform as well on math word problems that required relatively advanced reading skills, compared to word problems with text that was easier to understand. This pattern was observed for both the English Primary students and English Learners. However, in absolute terms, the weakest performance was consistently apparent for the ELs on the hard-text problems.

Hard text vs. easy text problems matched for math operation

The results outlined above pointed to the negative impact of challenging word problem text on students' math problem solving. However, it was possible that the word problems with difficult text also tended to involve more challenging mathematics. Therefore, we located word problems that involved the same math operation (e.g., addition) and then grouped these problems by text difficulty (Grade 7 or higher, under Grade 7 readability). Results for addition problems are presented next in some detail to illustrate the approach, followed by a summary of the results for other math operations.

Addition problems. There were 2,319 addition problem records. Table 6 includes the mean time (in seconds) for the easy-text and hard-text addition problems.

	Easy text	Hard text
Eng Learners	90.57 (104.73)	114.75 (143.21)
English Primary	86.33 (103.64)	92.19 (114.88)

Table 6. Mean time (seconds) per addition problem. Standard deviations are shown in parentheses.

An analysis of variance with language group (EL, EP) and text difficulty (easy, hard) as grouping factors indicated that students took longer on average on hard-text problems, $F(1,2318) = 8.433$, $p < .001$; ELs took longer than EPs, $F(1,2318) = 5.437$, $p < .05$; and the impact of hard text was greater for ELs, $F(1,2318) = 3.484$, $p < .06$.

Table 7 shows the mean number of incorrect answer attempts on each addition problem. ELs made more incorrect answer attempts than EPs, $F(1,2318)$, $p < .001$. No other effects were significant.

	Easy text	Hard text
Eng Learners	1.63 (1.28)	1.75 (1.28)
English Primary	1.41 (1.10)	1.49 (0.94)

Table 7. Mean number of incorrect answer attempts per addition problem. Standard deviations in parentheses.

Table 8 shows the mean activation of multimedia help per addition problem. ELs were more likely to activate the help features than EPs, $F(1,2318) = 22.866$, $p < .001$. There was no effect of text difficulty. The interaction of text difficulty and language status was suggestive, $F(1,2318) = 3.285$, $p < .07$.

	Easy text	Hard text
Eng Learners	0.08 (.27)	0.11 (.31)
English Primary	0.04 (.21)	0.03 (.19)

Table 8. Mean help activation per addition problem. Standard deviations shown in parentheses.

Table 9 shows the proportion of addition problems on which the correct answer was entered. An analysis of variance with language group (EL, EP) and text difficulty (easy, hard) as grouping factors indicated that EPs were more likely to answer correctly, $F(1,2318) = 23.312$, $p < .001$, and that performance was lower on hard-text problems, $F(1,2318) = 17.266$, $p < .001$. Also, the detrimental effect of hard text was significantly greater for the ELs than the EPs, $F(1,2318) = 8.087$, $p < .01$.

	Easy text	Hard text
Eng Learners	0.83 (.36)	0.72 (.44)
English Primary	0.88 (.32)	0.85 (.35)

Table 9. Proportion of addition problems with correct answers entered. Standard deviations in parentheses.

To summarize, the detailed analysis of addition problems indicated that both ELs and EPs performed less well with the hard-text word problems, meaning that they took longer on average on each problem and were less likely to find the correct solution than when the problem text was easy. In addition, there were some indications that hard text might have had a greater negative impact on the ELs than the EPs, including a reduced probability of ever finding the correct answer on a problem. Although students rarely activated the multimedia help features that were available in the problems, the ELs did so more often than the EPs and, in absolute terms, they were most likely to do so on the problems with challenging text.

Other math operations. We conducted similar analyses on problems involving subtraction (1313 problem records), multiplication (730 problem records) and division (817 problem records). Overall, the pattern of findings was very similar to that reported above for addition: ELs generally took longer on each problem and were less likely to enter the correct answer than EPs. Additionally, students performed less well on hard text problems than on comparable (in terms of the math operation) problems written in easy text. There were also consistent hints that the ELs were more affected by hard text than the EPs. For example, on subtraction problems, the ELs activated the help significantly more often on the hard text problems than the easy text problems, whereas there was no effect of text for the EPs. In no case did the ELs perform better on easy text problems than hard text problems.

Problem solving behaviors

The analyses described above focus primarily on the problem solving *outcomes* for English Learners and English Primary students. In the next set of analyses, we investigated students' *behaviors* as they worked on the problems. The time spent per problem was chosen as the primary metric. Initial calculations were performed to determine the average time required for each problem, along with the standard deviation for that problem. Because we had already established that the ELs worked more slowly overall, these calculations were performed twice for each problem: Once for ELs and again for the EPs. Then we cycled through the records for each student, comparing the time he or she spent solving each problem with the mean for his or her language group on that problem, and classifying the problem record into one of four possibilities:

"Gaming": The student answered the problem incorrectly up to 4 times in less than 40 seconds (average of 10 seconds per try). The rapid and incorrect responses suggested that the student was probably not reading the problem and was not seriously attempting to solve it.

"Fast solving": The student answered the problem correctly in his or her first attempt, and within 15 seconds of the problem presentation. The 15 second limit was chosen because it was the average time to read a problem and enter an correct answer.

"Solving": This classification referred to problems that were answered in more than 15 seconds but less than a standard deviation of the time other students of the same group (ELs or EPs) took to answer the same problem.

"Trying hard": These were cases in which the student took significantly more time on a problem than others in the same language group (i.e., more than a standard deviation above the mean for the group).

The results of the problem classifications may be seen in Table 10. (4% of problems completed by the ELs and 7% by the EPs did not fall clearly into one of the four categories, and are not included in Table 10.)

	Gaming	Fast solving	Solved	Trying hard
English Learners	4%	8%	62%	26%
English Primary	3%	2%	69%	26%

Table 10. Classification of problems completed by English Learners and English Primary students.

As may be seen in Table 10, the students were generally either solving or trying hard to solve most of the problems; gaming was quite rare. Also, the patterns are generally similar for the two language groups (ELs and EPs). Overall, the classification suggested that students took the activity reasonably seriously and that they were spending time trying to solve the problems.

The next issue was whether the English Learners' problem solving efforts were paying off to the same extent as was the case for English Primary students, particularly when the math problems were challenging to read. Table 11 shows the number of word problems with Grade 7 or higher readability on which students' behaviors suggested they were really trying to solve the problems (i.e., they took significantly longer than the average time for that problem, relative to others in their language group). As may be seen in the table, the English Learners were less likely to find the correct answer in these cases. Thus, it appeared that the difficult word problem text reduced their probability of a successful problem outcome.

	"Really Trying"	Correct Answer	Incorrect Answer
English Learners	435 problems	216 (49.6%)	219 (50.4%)
English Primary	493 problems	300 (61%)	193 (39%)

Table 11. Word problems classified as involving extra effort by students by problem outcome (correct or incorrect answer).

Discussion

The project goal was to evaluate the performance of English Learners and English Primary students who used AnimalWatch, an algebra-readiness ITS, as part of their math class instruction. Prior research suggested that the challenges involved in understanding English text might contribute to ELs' weaker math achievement, but this issue had not yet been investigated in the case of online instruction. The challenge was to find problem solving cases that were comparable for the two groups even though their overall performance was different, and their instructional experience in the ITS was different as well. Answering the question required transforming the original data in several ways: First, we found the math word problem records of over 400 high school students; second, we amplified the records with new information about the readability of the word problems, and third, we located problem records that had been solved by at least 20 students in each group (EL and EP).

The results were clear and consistent: English Learners' performance in the ITS was not as strong as that of the English Primary students. In addition, the analyses revealed that students' problem solving was influenced not only by the mathematics required to solve the problem but also by the complexity of the problem text. Both English Learners and English Primary students performed less well when the word problems involved challenging text. However, the negative impact of difficult text was especially apparent for the ELs.

One possibility for ELs' lower overall performance in the ITS was that they simply did not try as hard as the EP students. One might readily imagine that students who are faced with instructional materials in English that are difficult for them to understand might exert less effort or quickly give up. However, the data indicate that, if anything, the ELs tried harder. More specifically, they took longer on average on individual problems than the EPs; if the ELs had been giving up, a shorter average time per problem would have been predicted. ELs were also more likely than EPs to use the multimedia help features in the ITS (although it should be noted that use of the ITS help features was very low overall). The multimedia help features included interactives and video lessons, which the ELs may have found especially accessible. In addition, the time-based analyses of students' behaviors suggested that substantial proportions of ELs were trying to solve the problems, even though they were not always as successful as English Primary students. Thus, it does not seem likely that the lower overall performance of the ELs was due to lower effort.

In summary, the analyses revealed systematic differences in the problem solving performance of English Learners and English Primary students. The results point to the importance of considering multiple factors in designing effective ITS instruction, including the ways in which language is used to present the materials. One promising possibility is to extend the intelligent problem selector algorithm to consider the readability of word problems in addition to the mathematics operation when selecting problems for students who are not yet proficient with English.

References

August, D. L., & Shanahan, T. (2006). Developing literacy in second language learners: Report of the National Literacy Panel on language minority youth. Erlbaum Associates.

Beal, C. R., Adams, N., & Cohen, P. R. 2010. Reading proficiency and mathematics problem solving by English Learners. Urban Education, 45, 58-74.

Beal, C. R., & Cohen, P. R. (2005). Computational methods for evaluating student and group learning histories in intelligent tutoring systems. Proceedings of the 12th International Conference on Artificial Intelligence and Education, IOS Press.

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

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

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. Journal of the Learning Sciences, 13, 129-164.

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.

Acknowledgments

The AnimalWatch project has been supported by grants from the National Science Foundation (DRL 0903441, HRD 0120809) and the Institute of Education Sciences (R305K050086, R305K090197). The views expressed here are not necessarily those of the funding agencies. We would like to thank the teachers, staff and students of the participating schools for their support. Special thanks are due to Jamie Callan at Carnegie Mellon University for facilitating our access to the REAP online portal used in adding readability metrics to the AnimalWatch problems, and to Wesley Kerr at the University of Arizona for his heroic assistance with data retrieval and processing.