The Impact of Cognitive and Non-Cognitive Text-Based Factors on Solving Mathematics Story Problems

Candace Walkington
Southern Methodist University
3011 University Blvd. Ste. 345
Dallas, TX, 75205
1-214-768-3072
cwalkington@smu.edu

Virginia Clinton
University of Wisconsin - Madison
1025 West Johnson Street
Madison, WI 53706
1-608-890-4259
vclinton@wisc.edu

Steven Ritter
Carnegie Learning, Inc.
437 Grant Street, Suite 918
Pittsburgh, PA 15219
1-888-851-7094 x122
sritter@carnegielearning.com

Mitchell Nathan
University of Wisconsin - Madison
1025 West Johnson Street
Madison, WI 53706
1-608-262-0831
mnathan@wisc.edu

Stephen E. Fancsali
Carnegie Learning, Inc.
437 Grant Street, Suite 918
Pittsburgh, PA 15219
1-888-851-7094 x219
sfancsali@carnegielearning.com

ABSTRACT

Intelligent tutoring systems (ITSs) that personalize instruction to individual learner background and preferences have emerged in K-16 classroom settings all over the world. In mathematics instruction, ITSs may be especially important for tracking mathematical skill development over time. However, recent research has pointed to the importance of text-based measures when solving mathematics word problems, suggesting that in order to accurately model the student it is important to understand how they respond to text characteristics. We investigate the impact of text-based factors (readability and problem topic) on the solving of mathematics story problems using a corpus of N = 3394 students working through an ITS for algebra, Cognitive Tutor Algebra. We leverage recent advances in computerized text-mining to automate fine-grained text analyses of many different word problems. We find that several elements of the text of mathematics word problems matter for performance – including the concreteness of the problem's topic, the length and conciseness of the story's text, and the words and phrases used.

Keywords

Intelligent tutoring system, readability, mathematics, word problems, personalization

1. INTRODUCTION

Since the 1980s, Intelligent Tutoring Systems (ITSs) have risen as an important instructional tool to support student learning in classrooms, especially in middle and high school. ITSs typically consist of at least three components: (1) the *domain model* of the appropriate steps needed to correctly solve each problem, (2) the *student model*, which captures the evolution of an individual student's cognitive states as they relate to the domain model, and (3) the *tutoring model* which selects tutor actions based on the domain model and student model [1]. It is through the construction of the student model and its contribution to the tutoring model that ITSs can enact *personalization* where they adapt to the needs and backgrounds of individual learners. Here we explore cognitive and non-cognitive factors related to how students react to and understand the text of mathematics story problems. We argue that these non-mathematical factors may be an important element to consider for an ITS in secondary mathematics. In particular, we provide evidence suggesting that both the students' reading level (a cognitive factor) and the students' interests, preferences, and motivational outlooks (non-cognitive factors) have the potential to influence how they respond to text-based mathematics problems situated in "real world" contexts.

Cognitive Tutor Algebra (CTA; [2]) is a prominent mathematics ITS used in many schools across the United States. CTA uses *model-tracing approaches* to relate student actions to the domain model and provides individualized error feedback. CTA also uses *knowledge-tracing approaches* to track students' learning from one problem to the next, using this information to identify the students' strengths and weakness in terms of production rules (i.e., knowledge components or skills). The software then uses this analysis to individualize the selection of problem tasks. However, missing from this tutoring model is a consideration of other non-mathematical characteristics of the story problem texts – including the *reading difficulty* of the text respective to students' reading ability and preferences, and the *real-world topic* of the text respective to students' interests and preferences.

For example, a learner presented with a mathematics word problem that is difficult to read – with high-level vocabulary, complex sentence structure, etc. - may lack the reading ability to appropriately comprehend that problem. This cognitive element of the problem's difficulty is not typically monitored by ITSs for mathematics learning. In addition, such a problem may inhibit the students' motivation – a non-cognitive factor. In particular, even if the learner is technically able to read the problem, they may be intimidated by the problem text, and request a hint instead of putting forth the effort of understanding the text of the problem. ITSs also do not typically monitor the learner motivation for reading and understanding text-based problems.

Another non-mathematical element of the text of mathematics story problems is the real world topic – whether the story is about working at a part-time job or harvesting a field of grain. The way in which students react to the topic of the story problem is also based on both cognitive and non-cognitive factors. Students may be unfamiliar with elements of the context that are important for fully comprehending the problem – for example, in a banking context, they may not know what "break even" means. In this way, they may lack the prior knowledge needed to interpret the story. Similarly, different real world topics may differ in the motivation they elicit from students – students may experience greater motivation when solving a problem about a familiar, interesting context than about a context they find boring or unfamiliar.

We next provide a theoretical framework that provides an explanation of how students comprehend story problems and how cognitive and non-cognitive factors may interact as they solve story problems.

2. THEORETICAL FRAMEWORK

2.1 Cognitive Factors

Nathan and colleagues [3] proposed a model of mathematics story problem solving where students navigate three levels of representation as they comprehend and solve story texts: (1) a *textbase* containing the propositional statements made in the story problem, (2) a *situation model*, a qualitative representation of the actions and events in the story, and (3) a *problem model*, containing the formal mathematical equations, variables, and operands. Because mathematics word problems are stated in verbal language (rather than mathematics notation), we hypothesize that the reading difficulty and topic of the problem matters for the construction of the situation model and its successful coordination with the problem model.

Various aspects of the reading difficulty, including *readability measures*, may be important in situation model construction. Readability measures often include the kinds of words used, the length of the story, and the structure of the sentences. These elements of the text's structure may make it more difficult to comprehend, especially for students with weaker reading skills.

Another aspect of reading difficulty is the *topic* of the problem – whether it is about, for example, farming or banking. Walkington and colleagues [4] proposed that story contexts that are related to topics that are familiar and accessible to students are easier for them to solve because these contexts can facilitate situation model construction because of their relatedness to learner prior knowledge. In related work [5], they also identified the prevalence of issues with verbal interpretation of mathematics story problems, finding that even high school students struggle to understand difficult vocabulary words and construct an accurate propositional textbase and situation model from a story problem's text.

2.2 Non-Cognitive Factors

An important precursor to students' motivation is their level of *interest* – defined as the state of engaging and the predisposition to re-engage with particular topics, ideas, or activities [6]. Two types of interest have been described in the literature. First, *situational interest* is an immediate, temporary state of heightened attention and affective engagement that stems from elements of a learning environment that are surprising, salient, evocative, challenging, personally relevant, etc. Situational interest can be *triggered* in response to a stimuli within a learning environment, and then may or may not become *maintained* over time [6]. A second type of interest is *individual interest* – learners' enduring predispositions to engage with certain activities or topics over time.

Elements of a story problem's text have the potential to both trigger and maintain situational interest. In particular, story problems that are accessible, easy to read, and situated within the topics and contexts that a particular learner finds relevant and interesting may trigger and maintain interest. In the other hand, difficult reading passages disconnected from a learner's experiences and interests may not trigger interest and may cause disengagement if interest has previously been triggered.

2.3 Research Purpose

If text-based measures like readability and problem topic matter for student performance, these might be important elements to add to future systems for personalized learning in mathematics. For example, an ITS might present weak readers with problems with simplified verbal language as these learners are initially mastering a new mathematical skill. As the student gains expertise with the mathematics by mastering skills, additional levels of verbal difficulty could be layered on by the ITS. Similarly, learners that lack motivation may be presented with story problems that are less intimidating to read and situated within their interests, with this support faded out over time. By neglecting to model this aspect of the user's experience in the ITS, the system may be generating inferences about learner knowledge states that are inaccurate.

3. LITERATURE REVIEW

3.1 The Impact of Reading Difficulty on Solving Mathematics Story Problems

Recent research has found that reading ability is especially important as students solve mathematics word problems [7]. Studies examining the association of reading difficulty of mathematics word problems and U.S. student performance on large-scale assessments has found that problems that use words with multiple meanings, complex verbs, and mathematics vocabulary words are more difficult [8]; the effect is especially pronounced for students who speak English as a second language [9]. A small study of students working in CTA found that extraneous text that provided a real world context for the problem, as well as references to concrete people, places, and things, were associated with less concentration and more confusion in the tutor [10]. However, a similar study found that the extraneous text was also associated with fewer unproductive "gaming the system" behaviors in the tutor [11]. Converging evidence suggests text characteristics relating to reading difficulty are important when solving mathematics word problems, but studies are needed that address which elements of reading difficulty are most important.

3.2 The Impact of Problem Topic on Solving Mathematics Story Problems

The topic of mathematics story problems also has an important relationship to students' prior knowledge and motivation. A study of high school students solving either standard story problems or story problems personalized to topics they were interested in (e.g., sports, video games, social networking) within one unit of CTA found that personalized stories were associated with higher performance. This performance gain was present in two tasks – labeling independent and dependent quantities given in algebra story problems, and writing algebraic expressions from the story scenarios [12]. It was hypothesized that during these two tasks, students are working closely with the problem text, constructing their situation model and coordinating it with a problem model. This study also found that students receiving problems in the context of their out-of-school interests were less likely to game the system – to exploit regularities in hints and feedback provided by CTA in order to avoid productive learning behaviors. Further, students who received personalization had stronger performance in future units where the problems were no longer personalized.

In a recent follow-up study [13], story problems in four units of CTA were personalized to topics students were interested in, and students solving personalized problems were compared to a control group solving normal problems. Results showed that personalized problems both triggered students' situational interest and enhanced students' individual interest for learning algebra. Personalization was associated with greater learning gains than a control condition only when the personalization was matched to deep features of the students' interest area. This was contrasted with personalization that was only matched surface features of the learners' interests – i.e., modifications to the problems that simply involved inserting familiar pop-culture words rather than considering how learners might actually use relationships between quantities in their everyday activities. Thus converging evidence points to the importance of considering the real world topic of mathematics story problems and its relationship to students' interests and experiences. However, more research is needed to determine which topics may be more or less likely to trigger and maintain students' interest.

3.3 Research Questions

In the present study, we investigate the relationship between readability and topic measures and student performance on mathematics story problems. We examine these issues within an ITS for Algebra I, Cognitive Tutor Algebra (CTA), that tracks student hint requests in addition to whether they get problems correct or incorrect. We investigate two research questions: (1) How are readability and topic measures associated with correct answers and hint requests when students label independent and dependent quantities in stories in CTA? (2) How are readability and topic measures associated with correct answers and hint requests when students write algebraic expressions from stories in CTA? Answers to these questions could inform the design of future ITSs for personalized instruction.

4. METHOD

Data from N = 3394 students with active CTA accounts were collected from 9 high schools and 1 middle school that were diverse in terms of their socio-economic, racial, and achievement background (Table 1). Data were collected for students solving 151 distinct word problems accross the first 8 units of CTA; later units were not included because many students did not advance beyond these units. We collapsed for all analyses (i.e., treat as identical) problems containing an identical story but using slightly different numbers. On average, each problem had been solved by 742 students (SD = 495). Each problem included a story scenario that outlined one or more linear functions within a real world situation (Figure 1). The student was asked to complete steps in which they identified the independent and dependent quantities in the story, wrote a linear algebraic expression for the story, and solved their expression for different x and y values; we consider only the first two skills.

CTA log data from students in the selected schools were uploaded to DataShop (pslcdatashop.web.cmu.edu), an online repository of detailed student interaction data. These logs contained information on whether the student got each problem correct, incorrect, or requested a hint on their first attempt; because requesting a hint is a distinct outcome, correct and incorrect are not completely repetitive measures. Thus, for each problem, we compiled the percentage of students who had gotten the problem correct on the first attempt, incorrect, or requested a hint. This percentage was our dependent measure in three distinct regression models.We analyzed the text of the introduction to each story problem (i.e., the initial text that gives the linear rate of change and intercept; see Figure 1) with the *Coh-Metrix* and *LIWC* text-mining programs. Coh-Metrix

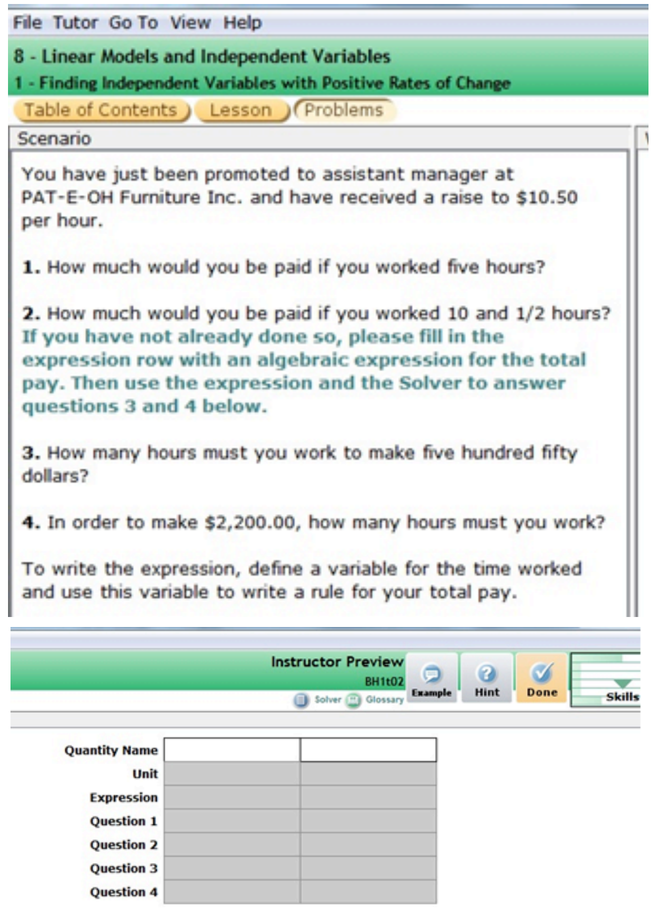
[14] measures a large number of aspects of text readability, including the amount semantic overlap between sentences, the number of verbs, use of concrete versus abstract words, the average sentence length, and others.

Table 1. Demographic characteristics of schools in study

ID	Math Prof %	State Prof %	School Enrollment	School Type
1	88%	70%	797	Middle
2	81%	47%	1,482	High
3	95%	84%	2,163	High
4	55%	46%	708	High
5	27%	NA	1,875	High
6	68%	59%	986	High
7	2%	31%	602	High
8	76%	84%	1,333	High
9	19%	39%	397	High
10	68%	79%	800	High

ID	White	Black	Hispanic	F/R Lunch
1	72%	7%	15%	21%
2	90%	4%	2%	4%
3	84%	10%	3%	6%
4	99%	1%	1%	41%
5	20%	4%	72%	77%
6	9%	2%	88%	41%
7	1%	99%	1%	82%
8	36%	60%	2%	48%
9	100%	0%	0%	45%
10	38%	51%	11%	62%

Because some of our story introductions had only one sentence, measures that pre-supposed multiple sentences were ommitted. LIWC [15] was used to determine the topic of the story problems – this program counts how many words in the story fall into various word categories, including social processes (family, friends, people), affective processes (positive emotions and negative emotions), biological processes (body, health, ingestion), cognitive processes (insight, causation, discrepancy, tentativeness, certainty, inhibition, inclusive/exclusiveness), perceptual processes (see, hear, feel), relativity processes (motion, space, time), and personal concerns (work, achievement, leisure, home, money, religion). If a story contained any words that fell into one of these topic categories, that story was coded as a 1 for that category; otherwise it was coded as a 0.



Answer Key:

Quantity Name	the time worked	the money earned
Unit	hour	dollar
Expression	X	10.5X
Question 1	5	52.5
Question 2	10.5	110.25
Question 3	52.381	550
Question 4	209.5238	2200

Figure 1. Screenshot of algebra story problem in CTA with answer key superimposed

For each category in Coh-Metrix and LIWC, the correlation was computed between the list of each problem's score on that category, and the percentage of students who got each problem correct, incorrect, or requested a hint. Correlations that were significantly different from 0 were tested for inclusion as fixed effects in regression models predicting the performance measures (hints, corrects, incorrects). These models included random effects that described various aspects of the problem's mathematical structure, including the unit and section it came from in CTA, and the numbers it used. Models were initially fit using the *lmer()* command in *R* including all potential fixed and random effects. Then we used the *step()* command in *R* to perform backwards elimination on fixed and random effects, leaving a model with only the effects that significantly improved the fit of the model. These analyses were carried out separately for a dataset that included only instances of students labeling independent and dependent quantities, and a dataset that included only instances of students writing algebraic expressions.

5. RESULTS

5.1 Labeling Independent and Dependent Variables

Regression results showing the relationship between performance measures (% incorrect, hint, and correct) and readability and topic measures for labeling quantities in story problems are provided in Table 2. Table 2 shows that problems that use adverbial phrases (*DRAP*) were associated with fewer incorrect answers. Adverbial phrases are phrases that add on to verbs, answering the questions where, when, or how? In the present data set, adverbial phrases mostly answered when the action occured, and often included words like *currently*, *already*, *next*, *first*, *every day/week*, and *not yet*. However, some of these adverbs also answered the how question, relying information about quantities that might be useful to cue students to the constraints of the problem – examples of words used in this manner included *only*, *completely*, and *evenly*. These words may have given important details about how the quantities involved in the story were changing as the action in the story proceeded.

Table 2. Regression tables relating performance measures on labeling quantities to readability/topic categories

	Estimate	Std. Err	t value	Pr(>|t|)	
% Incorrect					
(Intercept)	0.182	0.032	5.63	0.00018	***
DRAP	-0.0008	0.0003	-2.34	0.02104	*
motion	0.036	0.0137	2.65	0.00899	**
% Hint					
(Intercept)	0.045	0.014	3.21	0.01407	*
inhibition	0.023	0.008	2.83	0.00543	**
% Correct					
(Intercept)	0.784	0.044	17.87	0.00000	***
motion	-0.042	0.0182	-2.29	0.02370	*

Stories that involve *motion words* (e.g., *go*, *move*, *ran*, *arrive*, *come*, *enter*, *threw*) are associated with more incorrect answers and fewer correct answers. These stories often incldued contexts where people were walking, biking, hot-air-balooning, driving, or actively constructing something. In terms of the quantities used, there was often a rate of change (e.g., per hour, per minute, a day) that involved this motion, and students had to identify the two quantities that made up this rate of change. Using more abstract physics quantities – like distance and speed – may have been more difficult for students than using quantities relating to specific concrete objects (e.g., accumulating cards, toys, or money). Finally, *inhibition words* were associated with more hint requests. Inhibition words were often included in story problems that discussed safety issues or saving money. Students may have persieved these less concrete, finance- or safety-oriented contexts as less accessible, making them more likely to request a hint rather than attempt to write the labels. These problems often involved money as the dependent variable, but the label for this variable may have been complex because the actor in the story might have already saved or spent some money when the story started. Thus a label of simply *money* may not be appropriate, and the student would have to enter a label that captured that it was *total money* or *net money* saved or spent.

5.2 Writing the Algebraic Expression

Regression results showing the relationship between performance measures and readability and topic measures for writing the expression are shown in Table 3. We again see that *inhibition words* – often associated with financial contexts – are more difficult for students – they are associated with more incorrect answers, more hint requests, and fewer correct answers. The conceptual difficulty of this topic area might become especially important as students move from formulating their situation model to coordinating their situation model with a problem model.

Table 3. Regression tables relating performance measures on writing expressions to readability/topic categories

	Estimate	Std. Err	t value	Pr(>|t|)	
% Incorrect					
(Intercept)	0.195	0.060	3.26	0.00167	**
WRDPOLc	0.0494	0.013	3.91	0.00014	***
inhibition	0.086	0.034	2.52	0.01286	*
% Hint					
(Intercept)	0.055	0.014	3.95	0.00050	***
One sentence	(ref.)				
Two sentences	-0.045	0.016	-2.82	0.00548	**
Three Sentences	-0.057	0.017	-3.48	0.00067	***
4 + Sentences	-0.033	0.019	-1.77	0.07868	
RDL2	0.002	0.001	3.51	0.00061	***
family	0.030	0.015	2.05	0.04282	*
inhibition	0.052	0.011	4.74	0.00001	***
motion	0.025	0.009	2.77	0.00637	**
% Correct					
(Intercept)	0.334	0.17478	1.91	0.05778	
LDTTRc	0.428	0.169	2.53	0.01242	*
WRDPOLc	-0.041	0.01469	-2.78	0.00609	**

inhibition	-0.128	0.03909	-3.28	0.00132	**

Another factor that stands out in the regression results is word polysemy (*WRDPOLc*) – or the number of different meanings that a word has (for example, in English, *mine* can be something you own or an explosive device). The results show that stories that contain words with more potential meanings are associated with more incorrect answers and fewer correct answers. Polysemous words have been found to make mathematics word problems more difficult to interpret accross other studies [8-9].

Results also showed that higher type-token ratios (*LDTTRc*) are associated with more correct answers. As type-token ratio increases, more unqiue words are being used in the story problem, and fewer words are being repeated. These results suggest that students have an easier time writing the expression in a story that is relatively concise with little reptition of ideas. While it makes sense that this type of story may be more amenable to translation into mathematics notation, this result contrasts with research in text comprehension in reading tasks [14] which generally finds that repitition and lower type-token ratios facilitate reading comprehension. However, the story problems with high levels of word repetition frequently discuss complex topics of which students may lack familiarity, including operating capital, business inventory, and wholesale prices. In this way, a high type-token ratio may be indicative of a complex topic rather than increased readability in these story problems.

Students' tendency to seek hints when writing the algebraic expression is associated with a number of different readability factors. First, we see an effect for the length of the story text; students are more likely to seek hints for *one sentence story problems*, compared to problems that have two or more sentences. Having only one single sentence in a story problem might not be enough to ground or fully describe a linear rate of change as it arises in a real-world situation, and these overly-sparse stories might consequently inhibit performance.

In addition to greater difficulty of inhibition words, stories with *family words* and *motion words* were associated with greater hint-seeking. Only 13 of the problems involved family words, and these were often complex scenarios where multiple actors (e.g., a main character and his brother) were each contributing to the algebraic rate of change in their own way (e.g., saving/earning/splitting money together). Motion words often involved physics contexts (e.g., traveling in a car or plane) in which students had to track distance, rate, and time. This suggests that keeping track of multiple individuals engaging in mathematical actions and solving problems with physical distances and rates may be significant difficulty factors when writing expressions.

Finally, the regression results showed that scoring higher on Coh-Metrix's second language readability measure (*RDL2*) was associated with greater hint-seeking when writing expressions. This measure is calculated through measures of word frequency (with words that occur more frequently in the English language yielding higher scores), sentence syntax similarity (with sentences that have similar grammatical structures yielding higher scores), and word overlap (with words that share semantic meaning yielding higher scores; [16]). Given that a higher second language readability score is typically associated with greater ease in comprehending the text [17], it is suprising that stories that score higher on this measure would be associated with students seeking more hints. The explanation of this finding may be similar to that for our finding with type-token ratio; story problems that use similar words and sentence structures often use a lot of reptition as a way to present complex ideas. Stories that are simple and concise may be easier for students to solve.

6. DISCUSSION

Results indicate that readability and topic measures have important associations with students' performance when solving mathematics word problems in an ITS. In particular, it was more difficult for students to name the independent and dependent quanitities in problems relating to motion (physics) and inhibition (saving and safety), while adverbial cues facilitated this skill. When writing algebraic expressions, we again see that motion and inhibition topics are difficult, but also find other important readability measures that matter. Words with multiple meanings make story problems more difficult, which corresponds to previous findings in both mathematics and reading education.

However, mathematics stories that use concise language with little repitition, which in terms of their readability level makes them technically *less* readable, are actually easier for students to solve. Thus measures of readability that stem from research on reading comprehension may need to be considered differently when working with mathematics problems. Results also suggest that while a story problem that includes only a single sentence is concise, it might present difficulty for students by not providing necessary context and information for them to feel they can respond without needing a hint.

Overall, our results suggest that mathematics story problems that have story texts that are more accessible to students have several characteristics: (1) they are concise with little repetition, but not a single sentence only, (2) they use only a single actor performing actions, (3) they use simple words with clear meanings, (4) they avoid more abstract physics or financial contexts, instead focusing on familiar contexts involving accumulation or loss of concrete physical objects, and (5) they make use of adverbial cues. Story problems with these characteristics may allow students to more easily construct a situation model from a propositional textbase. They may promote situation-model construction by both increasing students' ability to comprehend the semantics of the problem, and by increasing students‘ interest in working on the problem.

7. CONCLUSION

Future adaptive ITSs will be designed to model student characteristics at an extremely fine-grained level, as technology for personalized learning continues to advance. Here we argue that an important element of these future adaptive systems will be a consideration of the non-mathematical text-based characteristics of the problem tasks they present to students. Making inferences about students‘ current level of mathematical knowledge or motivation without considering these characteristics may lead to misspecifications.

Readability and topic measures may be an important consideration for ITSs to model in a variety of domains, including when considering tasks from history, social studies, and science. Future research should focus on the readability and topic measures that are most important for students of different age groups in different subject domains, and narrow down which characteristics are most critical to include in student and domain models as we build future ITSs. In current work, we are analyzing the mathematics problems on the National Assessment of Educational Progress (NAEP) and Trends in International Mathematics and Science Study (TIMSS) to examine how readability and topic measures impact the performance of 4th and 8th graders in the United States, and how these factors interact with cognitive and non-cognitive student background characteristics.

8. REFERENCES

[1] Padayachee, I. 2002. Intelligent tutoring systems: Architecture and characteristics. University of Natal, Durban, Information Systems & Technology, School of Accounting & Finance.

[2] Ritter, S., Anderson, J. R., Koedinger, K. R., Corbett, A. 2007. Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14(2), 249-255.

[3] Nathan, M. J., Kintsch, W., Young, E.: A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, 9(4), 329-389 (1992)

[4] Walkington, C., Petrosino, A., Sherman, M. 2013. Supporting algebraic reasoning through personalized story scenarios: How situational understanding mediates performance and strategies. *Mathematical Thinking and Learning*, 15(2), 89-120.

[5] Walkington, C., Sherman, M., & Petrosino, A. 2012. ‘Playing the game’ of story problems: Coordinating situation-based reasoning with algebraic representation. *Journal of Mathematical Behavior*, 31(2), 174-195.

[6] Hidi, S., & Renninger, K. 2006. The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111-127.

[7] Vilenius-Tuohimaa, P. M., Aunola, K., Nurmi, J. E. 2008. The association between mathematical word problems and reading comprehension. *Educational Psychology*, 28(4), 409-426.

[8] Shaftel, J., Belton-Kocher, E., Glasnapp, D., Poggio, J. 2006. The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*, 11(2), 105-126.

[9] Wolf, M. K., Leon, S. 2009. An investigation of the language demands in content assessments for English language learners. *Educational Assessment*, 14(3-4), 139-159.

[10] Doddannara, L. S., Gowda, S. M., Baker, R. S., Gowda, S. M., De Carvalho, A. M. 2011. Exploring the relationships between design, students’ affective states, and disengaged behaviors within an ITS. In: *Proceedings of the 16th International Conference on Artificial Intelligence and Education*, pp. 31-40.

[11] Baker, R. S., de Carvalho, A. M. J. A., Raspat, J., Aleven, V., Corbett, A. T., Koedinger, K. R. 2009. Educational software features that encourage and discourage “gaming the system.” In: *Proceedings of the 14th International Conference on Artificial Intelligence in Education*, pp. 475-482.

[12] Walkington, C. 2013. Using learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology*, 105(4), 932-945.

[13] Bernacki, M. & Walkington, C. 2014. The Impact of a Personalization Intervention for Mathematics on Learning and Non-Cognitive Factors. Submitted to the *2014 International Conference of Educational Data Mining*, London.

[14] Graesser, A. C., McNamara, D. S., Louwerse, M. M., Cai, Z.: Coh-Metrix: Analysis of text on cohesion and language. Behavior Research Methods, Instruments, & Computers, 36(2), 193-202 (2004)

[15] Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., Booth, R. J.: The development and psychometric properties of LIWC2007. Austin, TX, LIWC. Net. (2007)

[16] Crossley, S., Allen, D., McNamara, D. 2011. Text readability and intuitive simplification: A comparison of readability formulas. Reading in a Foreign Language, 23(1), 84-101.

[17] Crossley, S. A., Greenfield, J., McNamara, D. S. 2008. Assessing text readability using cognitively based indices. TESOL Quarterly, 42(3), 475-493.