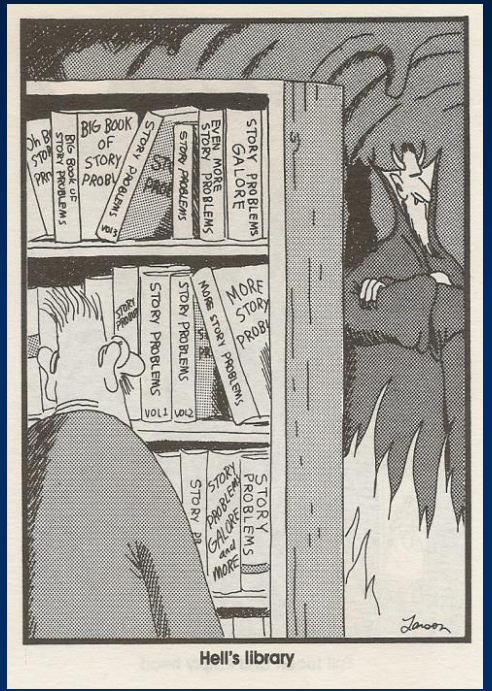
The Impact of Cognitive and Non-Cognitive Text-Based Factors on Solving Mathematics Story Problems

Candace Walkington - Southern Methodist University
Virginia Clinton, University of Wisconsin – Madison
Steve Ritter – Carnegie Learning Inc.
Mitchell Nathan – University of Wisconsin - Madison
Steve Fancsali – Carnegie Learning Inc.

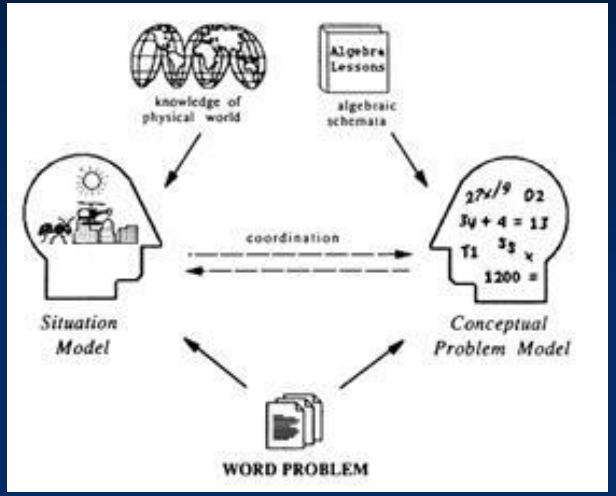
Introduction

- Student **motivation** and **interest** important non-cognitive components of performance in tutoring environments
- Finding ways to **activate** and **sustain** the development of interest has the potential to enhance students' learning in these environments (Walkington & Bernacki, in press)
- Different students have different needs, dependent upon **prior knowledge** (Mayer, 2001) and **attitudes** (Walkington et al., 2013)
- A better understanding of how cognitive and non-cognitive factors interact as students solve problems can offer guidance for efforts to **personalize** learning



Comprehending Story Texts

- In order to solve story problems, learners must comprehend problem text (Kintsch & Greeno, 1985)
- Successful problem-solvers coordinate:
  - ***Surface model***: Literal text
  - ***Textbase***: Info arranged in propositional form
  - ***Situation model:*** Understanding of actions and relationships
  - ***Problem model:*** Formal equations and operands
- The *readability level* and *topic* of problems may interact with cognitive and non-cognitive factors



Nathan, Kintsch, & Young (1992)

Theoretical Framework – Cognitive Factors

- **Cognitive Load** (Sweller, Merrienboer, & Paas, 1998):
  - A story that is easier to read may reduce extraneous cognitive load, freeing up resources for germane processing
  - A story that is familiar may provide grounding (Koedinger et al., 2008):
    - Grounded representations are more easily accessed in long-term memory
    - Less prone to errors given redundant semantic elaborations in long-term memory that can be used to support or check inferences
- **Individual differences principle** (Mayer, 2001):
  - Design effects intended to reduce cognitive load more important for low-knowledge learners because high-knowledge learners can better compensate for less support

Theoretical Framework – Non-Cognitive Factors

- **Situational Interest**
  - Spontaneous and transitory reaction to features of environment
  - Can focus attention, promote persistence, and improve learning (Ainley, Hillman, & Hidi, 2002; Harackiewicz et al, 2008)
  - Text characteristics associated with situational interest include: coherence, completeness, informational complexity, concreteness, ease of comprehension, imageability, suspense, importance/relevance of information, and identification with characters (Schraw & Lehman, 2001)
- Texts that are **more readable** (i.e., more coherent and less complex) may activate situational interest, depending on background of learner
- Texts that are connected to **personally relevant topics** may also activate situational interest
- Effects of situational interest interventions vary based on attitudes towards math (e.g., Bernacki & Walkington, 2014; Durik & Harackiewicz, 2006)

Readability of Story Problems

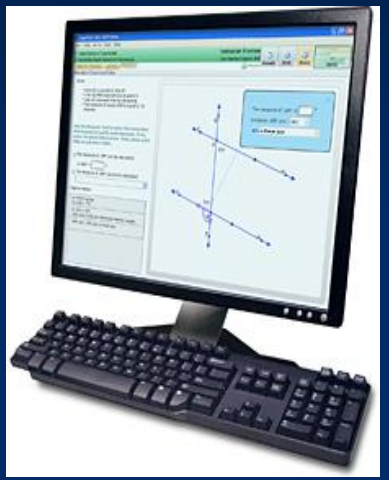
- Research has shown that reading skill is a strong predictor of performance on math story problems (Kyattla & Bjorn, 2013)
- Research has not successfully linked traditional readability measures (words, syllables, etc.) to mathematics performance (Wiest, 2003)
- Research on text comprehension outside of mathematics suggests that many characteristics influence readability at the surface (wording and syntax), textbase (explicit ideas), and situation model (meaningful representation) levels (Graesser & McNamara, 2011)

Topics of Story Problems

- Some problem topics may be more or less familiar to students
  - Urban students in the South solving problems about shoveling snow or building a greenhouse (Walkington, 2010)
- Matching problem topics to students' interests and experiences (i.e., **context personalization**) can improve both immediate performance and long term learning in mathematics (Walkington, 2013)

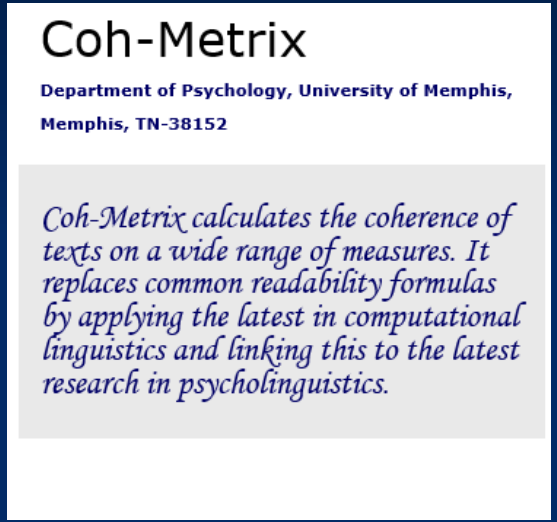
Research Questions

- **RQ1**: How are quantitative measures of **text readability** associated with performance on math story problems?
- **RQ2**: How are quantitative measures of **topic incidence** associated with performance on math story problems?
- **Two contexts**: ITS and standardized tests



Readability: Coh-Metrix

- Provides 108 quantitative measures of (McNamara et al., in-press):
  - **Surface code**: Difficulty of words and syntax
  - **Textbase**: Ease of connecting different ideas in the text
  - **Situation model**: Consistency of various dimensions of the mental representation of the text such as causation, time, and space



Topic Incidence - LiWC

- Dictionary-based text analysis program (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007)
- The topic of a story problem was determined by whether it had at least one word that fell in one of the LIWC topics or sub-topics

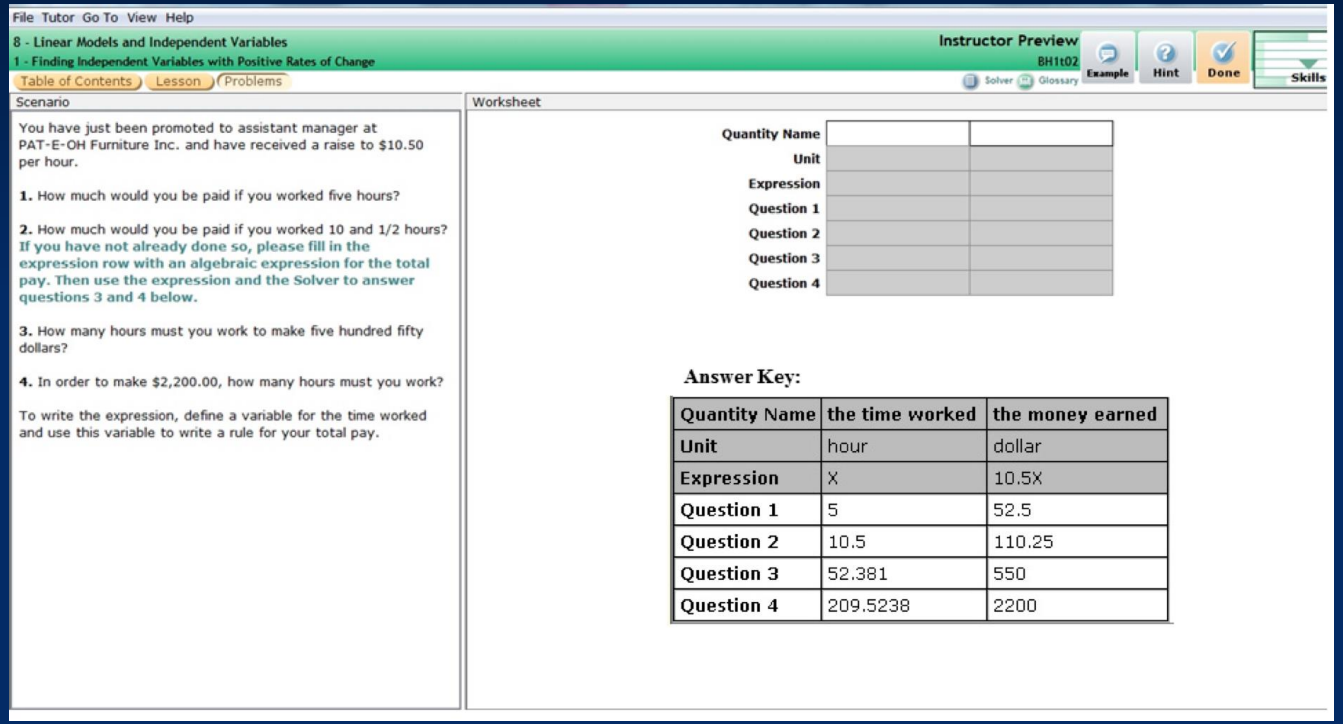
LIWC Topics	LIWC Sub-Topics
Social	family, friends, people
Affective	positive emotions and negative emotions
Biological	body, health, ingestion
Cognitive	insight, causation, discrepancy, tentativeness, certainty, inhibition, in/exclusiveness
Perceptual	see, hear, feel
Relativity	motion, space, time
Personal	work, achievement, leisure, home, money, religion, death

Study 1: Participants

- 9 high schools and 1 middle school using Cognitive Tutor Algebra (CTA) curriculum (*N* = 3394 students)

School Characteristic	Schools in Sample
Geographic Location	10 different states, 2 rural, 4 urban, and 2 suburban
F/R Lunch %	3 had 0-33% of students eligible, 4 had 33-66% eligible, 2 had 66-100% eligible
Race/Ethnicity	5 predominately White , 3 predominantly African-American, 2 predominately Hispanic
Achievement Level	3 had under 50% of students proficient, and 3 had 80% or more proficient, 4 in between

Study 1: Environment



Data collected from the first 8 units in CTA on linear functions – 151 unique story problems

Study 1: Analysis Techniques

- Logistic regression models predicting: **Correct**, **Hint**, **Incorrect**
- (Linear regression model predicting: **Latency**)
- Tested only readability/topic indicators with significant correlations to one or more of outcomes
- Sample size was ***N* = 151 problems** - each problem's performance measures were averaged for all students who solved that problem
- Random effects to control for mathematical aspects of problem, as well as for student effects
- Predictors tested for inclusion using Chi-Square test for significant reductions in deviance of the model
- **Three models**: All problems, expression-writing, and low schools

Study 1: Results

- **Moving from 3 to 4 sentences** reduced corrects by an estimated 4.4% ($p < .01$), increased incorrects by an estimated 2.75% ($p < .05$) and increased hints by an estimated 1.4% ($p < .01$)
  - Other measures of length (DESWC, DESSL) sometimes significant as well.
- **Third person singular pronouns** (3PS; he/she/it) associated with more correct answers ($p < .05$) and fewer hints ($p < .01$)
  - Changing a problem with no 3PS to a problem that has 10% of its words as 3PS should increase correct answers by an estimated 3.6%
- Increasing the **standard deviation of the amount of semantic overlap between sentences** significantly decreased correct answers
  - Moving from SD = 0 to SD = 0.3 would decrease corrects by 4%
- Using **concrete words** significantly reduced hints *(low schools only)*
- Using **words with multiple meanings** (word polysemy) significantly decreased corrects and increased incorrects *(expression-writing only)*
  - E*ach* additional meaning decreases corrects by 4.6% and increases incorrects by 4.6%

Study 1: Results

Category	High Example	Low Example
Standard deviation of overlap of adjacent sentences	Ms. Williamson woke up one morning to find her basement flooded with water. She called two different plumbers to get their rates. The first plumber charges $75 just to walk in the door plus $25 an hour. The second plumber charges a flat $40 an hour. (*SD* of overlap = 0.375)	You have just become CEO (Chief Executive Officer) of a company that is heavily in debt. The company's balance sheet currently shows a balance of -$525000. The company is paying the debt off at the rate of $12500 per month. (SD of overlap = 0.025)
Word polysemy	An open **pit** copper **mine** is 1550 feet deep and the company estimates that it is getting deeper at the rate of seven **feet** per month. Assume the number of feet below the surface is a negative number. (Average number of meanings of each content word = 6.375)	On Tuesday morning at 7 AM the residents of Bar Harbor Maine awoke to six inches of snow on the ground. The snow fell at the average rate of one-half inch per hour during the storm. (Average number of meanings of each content word = 2.342)

Study 1: Results

- Words involving **social processes** decrease incorrect answers (p<.05) by an estimated 2.09%
  - References to family, friends, humans, socializing, speaking, having parties, making calls, sending messages
- **Work words** decrease corrects (p < .05) by an estimated 1.89%
- **Motion words** decrease corrects (p < .05) by an estimated 2.03%
- **Healthcare words** increase hints (p < .01) by an estimated 1.63%
- Inhibition words (**saving money**) increase hints (p < .05) by an estimated 0.98%

Study 1: Results

Topic	Example Problem
Social	A bride is making nameplates to put on the tables at her reception. She can make them at the rate of 25 per hour. She works for two hours and quits for the night realizing that she cannot complete this many nameplates herself. The next day she calls her mother and they both work together. Her mother can make 35 nameplates per hour.
Work	You have just been promoted to assistant manager at PAT-E-OH Furniture Inc. and have received a raise to $10.50 per hour.
Motion	A machine called the Crawler which moves space shuttles travels at the rate of 29 feet per second. The Crawler is currently 100 feet from the hanger moving toward the launching pad.
Health	According to the American Heart Association approximately 145000 women die every year from smoking-related diseases. In fact lung cancer has become the leading cause of cancer death among women.
Inhibition	During the school year teachers save money for use during the summer when they're not being paid. This year due to some unexpected expenses one teacher was able to save only $879. He figures he will need $23 a day for personal spending money.

Study 1: Replication

- Eighth grade students solving *N* = 60 MATHia problems on linear functions
- **Coh-Metrix:**
  - Results replicate, except **length** (limited problem set) and **word polysemy**
- **LIWC:**
  - Not enough non-social contexts to replicate that finding, but **home** references associated with higher performance
  - **Work** contexts associated with lower performance
  - **Motion** finding does not replicate

Study 1: Discussion

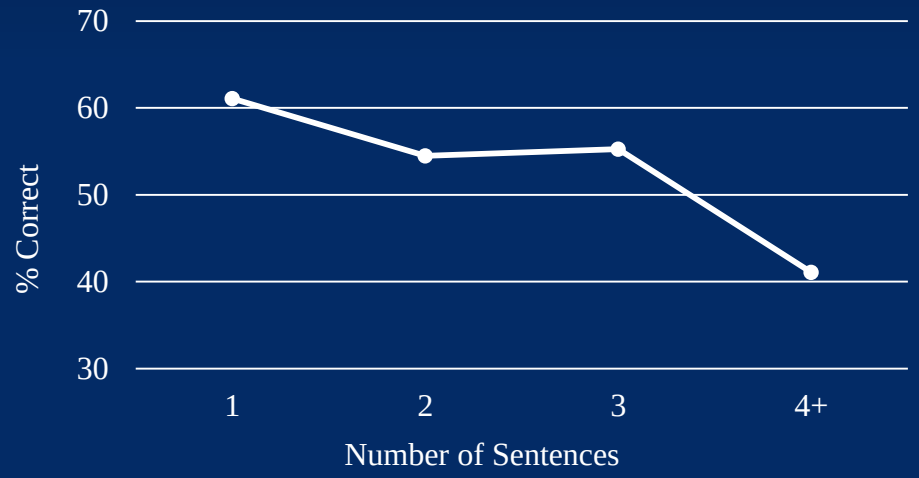
- **Good** for problem solving performance: Shorter stories, concrete words, third person singular pronouns, social contexts
- **Bad** for problem solving performance: Longer stories, stories with inconsistent overlap, words with multiple meanings, healthcare, business, financial, or physics contexts
- **Effect sizes are small** – ΔR^2-like measures fall between 5 and 15% typically; individual correlations around 0.2
- Makes sense that ***student*** characteristics and ***mathematical*** characteristics of the problem should explain most of performance

Study 2: Design

- Conducted Coh-Metrix/LIWC analyses on 700 released 4th and 8th grade NAEP mathematics items
- ~2 million U.S. students have taken the NAEP since 1996
- Looked at only items that had at least 1 sentence worth of words; predictors to control for:
  - Presence of visuals in the question or in the answer choices
  - Multiple choice or short constructed response
  - Grade level
  - Mathematical topic area
  - Difficulty (Easy, Medium, Hard) and Complexity (Procedural, Conceptual, Problem-solving)
  - Whether answer choices were words or numbers
  - Whether problem was in a “real world” context
  - Availability of calculator and manipulatives

Study 2: Results

- Readability/Topic measures may be less important on standardized tests
- However, length still important:



Study 2: Results

- Readability measures:
  - More sentences (4+) associated with lower performance
  - Longer sentences associated with lower performance *(4th grade only)*
  - Second-person pronouns associated with lower performance
  - Other pronouns associated with higher performance
  - Word acquired later in life associated with lower performance
- Topic measures:
  - Insight words (think, know, consider) associated with lower performance
  - Leisure words associated with higher performance *(4th grade only)*
  - Problems in real world contexts were easier than other word problems
  - Current: Recoding topics of problem topics by hand

Upcoming Analyses

- Look at how NAEP study results vary by student background characteristics and problem characteristics
- Conduct similar analyses with TIMSS released items
- Look at which readability/topic factors predict differences in TIMSS item performance between U.S. and other groups of countries

Discussion

- Within a mathematics curriculum, measures relating to surface model (word difficulty), textbase (length, overlap, pronouns), and situation model (topics) are all associated with performance
- Within a standardized math test, only surface model (word difficulty) and textbase (length, pronouns) measures seem to have significance
- **Cognitive load** may be of paramount concern on standardized tests, with little opportunity to elicit or maintain **situational interest** through interesting problem topics and concrete, relevant language

Implications

- In algebra, need to learn to navigate complex and potentially lengthy story contexts on diverse array of real world/professional topics
- To introduce and provide access to **new, challenging ideas**, readability and topic findings can be taken into account by problem designers
- Personalization/adaptation to learner needs:
  - **Struggling with a concept**: Simple, readable problems in interesting and accessible real world contexts
  - **Some mastery of concept**: Some factors of difficulty relating to readability present in most problems; mix of interesting and dry/application contexts
  - **Full mastery of concept**: Completely abstract/symbolic problems or challenging story problems with high-level language in contexts of authentic application

Next Steps

- Intervene on readability measures – all findings are strictly correlational
  - *ASSISTments* platform?
- Several studies show intervening on topic measures is effective for promoting interest and learning (Walkington, 2013; Bernacki & Walkington, 2014), with additional studies underway

Acknowledgements

Thank you to **Mitchell Nathan** and **Steve Ritter** for their support of this project!

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305C100024 and R305B100007. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

This research was also supported through a grant from the American Educational Research Association, and through the continuing collaboration of Carnegie Learning Inc.

Questions? Comments?