





Improving Comprehension Strategies of Struggling Adult Readers through Conversational Trialogues with AutoTutor

Art Graesser











MEMPHIS.

Overview

- Conversational Agents in Learning Environments
- Dialogues, Trialogues, and N-alogs
- AutoTutor Trialogues to Help Struggling Adult Readers

Why focus on adults with low literacy?

- 1 out of 6 adults in the US do not read well enough for them to get a decent job (National Research Council, 2011; Programme for International Assessment of Adult Competencies, OECD, 2011).
- Attendance is a problem because work schedules, childcare issues, and transportation difficulties (Greenberg, Reder, Rosen).
- Comprehension training interventions for adult readers are few in number, with weak evidence they are helpful (Greenberg, Mellard, Sabatini).
- Al technology can come to the rescue by improving comprehension training and providing intelligent support, 24-7!

Graesser, A.C., Greenberg, D., Olney, A.M., & Lovett, M.W. (in press). Educational technologies that support reading comprehension for adults who have low literacy skills. In D. Perin (Ed). *Wiley adult literacy handbook*. New York: Wiley.

Organizations committed to improve adult literacy

Commission on Adult Basic Education: http://www.coabe.org/

Institute of Education Sciences: http://ies.ed.gov/

Literacy Information and Communication System: http://lincs.ed.gov/

Office of Career, Technical, and Adult Education:

http://www2.ed.gov/about/offices/list/ovae/index.html

ProLiteracy: http://www.proliteracy.org/

VALUEUSA: http://www.valueusa.org/

Daphne Greenberg
Jan Frijters
Art Graesser
Maureen Lovett
Lee Branum-Martin
Andrew Olney



Enter here











The research reported here is supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C120001

Georgia State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Conversational Agents in Learning Environments

- Graesser, A.C., Rus, V., Hu, X. (2017). Instruction based on tutoring. In R.E. Mayer and P.A. Alexander (Eds.), Handbook of Research on Learning and Instruction (pp. 460-482). New York: Routledge Press.
- Nye, B.D., Graesser, A.C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, 24, 427–469.

Adaptive Intelligent Conversational Agents



STEVE



Guru (biology)



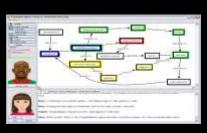
DeepTutor (physics)



AutoTutor Trialogs



iSTART (reading)



Betty's Brain



AutoTutor Trialogs with ALEKS algebra



Herman-the-Bug



Tactical Language and Culture System



Mission Rehearsal

Memphis Intelligent Conversational Agents



AutoTutor (computer literacy)



Guru (biology)



Which is one of the specific cases at the drug?

Charles and the specific cases at the drug?

Charles are considered as the specific cases at the specific case at the specific case

AutoTutor (reading comprehension)



AutoTutor (with ALEKS algebra)



ARIES (scientific reasoning)



HURA Advisor (research ethics)



ElectronixTutor



Personal Assistant for Lifelong Learning (PAL3)

Functions of Conversational Agents

- Help when initiated by the user
- Navigational guide
- Modeling action, thought, and social interaction
- Adaptive intelligent conversational dialog
- Staging arguments to prompt deeper learning
- Staging scenarios for assessment
- Many roles: peers, tutor, mentor

Emotions During Learning

(Graesser, Baker, Craig, D'Mello, Lehman, Rodrigo)

Boredom (23%)



Confusion (25%)



Delight (4%)









Frustration (16%)



Surprise (4%)

Measures collected at different grain sizes

- 1) Lessons attempted and completed
- 2) Performance in each lesson
- 3) Selecting answers to multiple choice questions
- 4) Semantic matches between natural language input and expectations or misconceptions
- 5) Initiative by asking questions, selecting tasks, and performing unprompted actions
- 6) Fluency of language and action
- 7) Engagement by response time patterns & coupling with item difficulty
- 8) Emotions (confusion, frustration, boredom, etc.)

Meta-analyses on Intelligent Tutoring Systems

		Effect Size
Kulik & Fletcher (2016)	50 comparisons	0.66
VanLehn (2011)	54 comparisons, STEM	0.58
Ma, Adesope, Nesbit, & Liu (2014)	107 comparisons	0.43
Steenbergen-Hu & Cooper (2014)	39 comparisons, college	0.35
Steenbergen-Hu & Cooper (2013)	26 comparisons, math, K12	0.05
Ritter, Kulikowich, Lei, et al. 2007)	Cognitive Tutor, math, WWCH	0.38
Fletcher & Morrison (2012)	Digital Tutor (1-study, N = 26)	3.17
Nye, Graesser, & Hu (2014), Graesser (2016)	AutoTutor (science, dozens of studies)	0.60 to 0.80

Generalized Intelligent Framework for Tutoring

Army Research Lab and University of Memphis

www.gifttutoring.org

Sottilare, R., Graesser, A., Hu, X., & XXXX (2013-2018). Design Recommendations for Intelligent Tutoring Systems.

- Learner modeling (2013)
- Instructional strategies (2014)
- Authoring tools (2015)
- Domain knowledge (2016)
- Assessment (2017)
- Teams (2018)
- Self-improving systems (2019)





Institute for Intelligent Systems

THE UNIVERSITY OF MEMPHIS

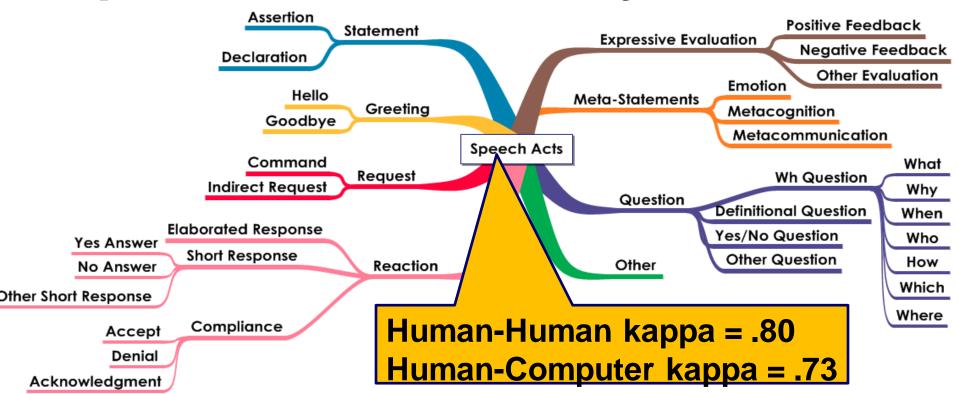
Dialogues and Trialogues

- Graesser, A.C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education*, 26.124-132.
- Graesser, A.C., Forsyth, C., & Lehman, B. (2017). Two heads are better than one: Learning from agents in conversational trialogues. *Teachers College Record*, 119, 1-20.

Expectation & Misconception-Tailored Dialog

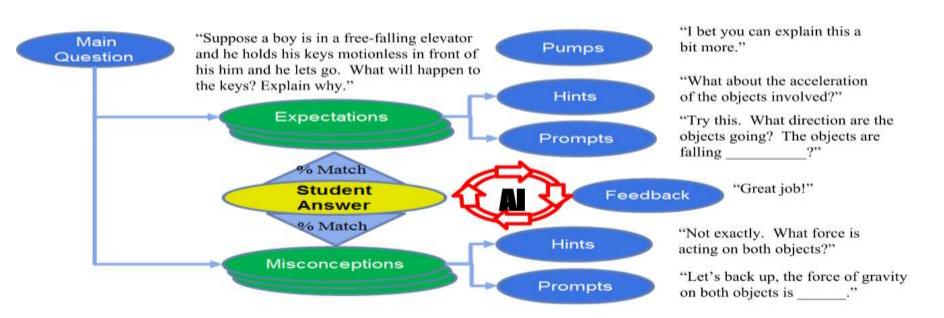
- Tutor asks question that requires explanatory reasoning
- Student answers with fragments of information, distributed over multiple turns
- Tutor analyzes the fragments of the explanation
 - Compares to a list of expected good idea units (via LSA and Regular expressions)
 - Compares to a list of expected errors and misconceptions
- Tutor posts goals & performs dialog acts to improve explanation
 - Fills in missing expected good idea units (one at a time)
 - Corrects expected errors & misconceptions (immediately)
- Tutor handles periodic sub-dialogues
 - Student questions
 - Student meta-communicative acts (e.g., What did you say?)

Speech Act Hierarchy



AutoTutor-Style (EMT) Dialog

Dialog with student cover expectations & correct misconceptions



Managing One AutoTutor Turn

Short feedback on the student's previous turn

Positive feedback: "Yeah" "Right!" Neutral feedback: "Okay" "Uh huh" Negative feedback: "No" "Not quite"

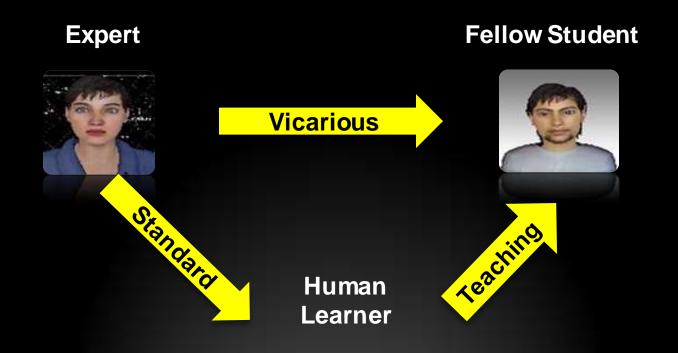
- Advance the dialog by one or more dialog moves that are connected by discourse markers
- End turn with a signal that transfers the floor to the student
 - Question
 - Prompting hand gesture
 - Head/gaze signal

Challenges

- Semantic matches are pretty good but not perfect

 Understanding is limited
- Semantic blur between expectations and misconceptions
 Some errors in feedback
- Learners expect full credit when they express a couple of juicy words instead of a sufficiently articulated statement ->
 The can be irritation or frustration
- High verbal or knowledgeable learners read printed conversation faster than listening to an agent
- Limited ability to handle student questions and requests

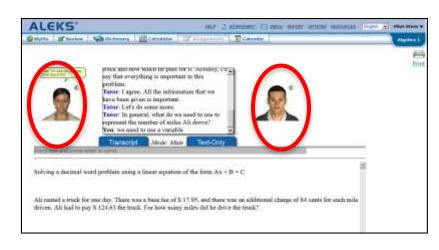
Adaptive Trialogues



Trialogues in Literacy and Numeracy

Xiangen Hu, University of Memphis











Advantages of Trialogues

- Two agents can model desired social interactions
- Two agents can disagree, stage an argument, and create cognitive disequilibrium
- A peer agent can echo a learner's contribution in a well articulated language, so:
 - Agent gets blame for a bad answer
 - Agent and learner gets credit for a good answer
 - Learner sees a well articulated response.
- Peer agent model good inquiry and receive good responses from the tutor agent



Trialogue (English Language Skills)

Utterance
Hey, Ron, you need to leave your water outside. I'm going to go talk to my friends. I'll see you guys inside.
Why did she tell me I have to leave my water outside, Tim?
I don't know.
Tim, why can't I drink water?
The books may get wet.
Why do you still have your water bottle, Ron? Look at rule number 2. We cannot get in the library with food or drink.

AutoTutor Trialogues to Help Struggling Adult Readers



Enter here







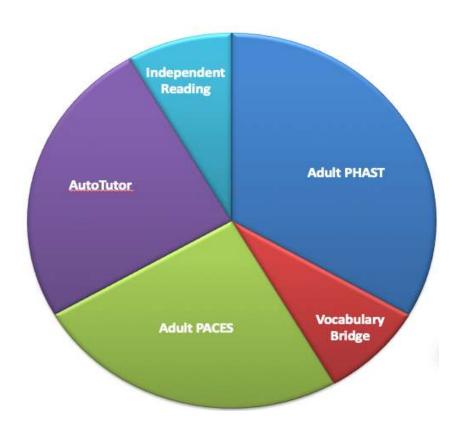




The research reported here is supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C120001

Georgia State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Intervention to Improve Reading for Adults



- Adults who read at grade levels 3-7.9
- 100 instructional hours
- · 2 times weekly
- 2-3 hours per duration
- Atlanta and Toronto areas
- 253 in the intervention
- Comparison to business as usual

Adult PACES Comprehension Program

- P = Predicting purpose with text signals and key information.
- A = Acquiring vocabulary with context clues.
- C = Clarifying common sources of confusion with clarifying questions.
- E = Evaluating and elaborating through questioning.
- S = Summarizing with text maps.

Lovett, M.W., Lacerenza, L., De Palma, M., & Frijters, J.C. (2012). Evaluating the efficacy of remediation for struggling readers in high school. *Journal of Learning Disabilities*. 45, 151-169.

Multilevel theoretical framework of discourse comprehension

Words

Syntax

Textbase

Explicit ideas (propositions)
Referential cohesion

Situation model

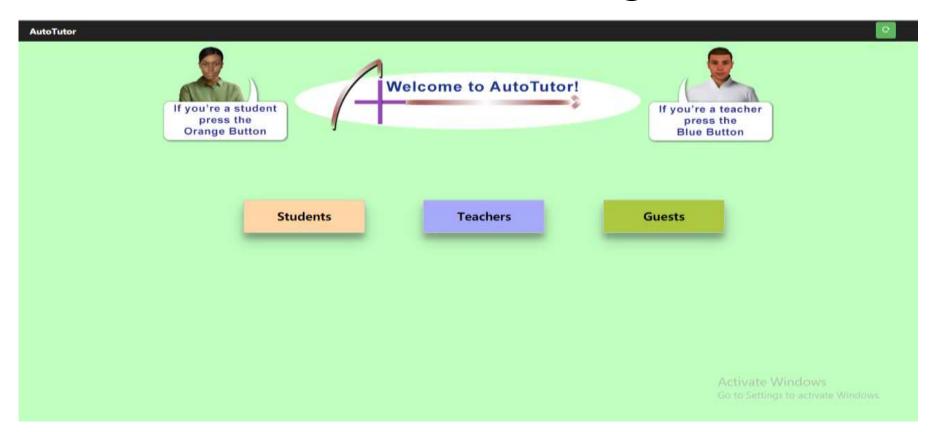
Causal, intentional, temporal, spatial, logical relationships Connectives, signaling words

Genre and rhetorical structure Pragmatic communication

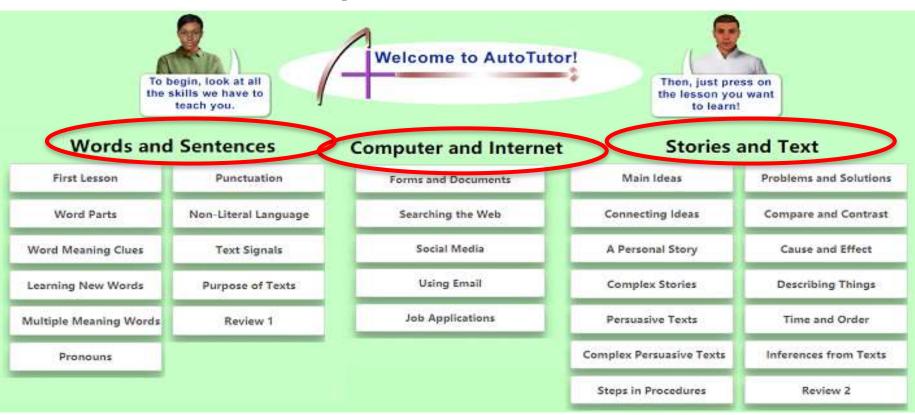
Graesser, A.C., & McNamara, D.S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3, 371-398.

Also Goldman, Kintsch, Perfetti...

Home Web Page



Scope of Lessons



Database for Instructor



More Details about AutoTutor for CSAL

35 lessons on comprehension

20-60 minutes each
Summary Nugget → Conversational Training
Cover theoretical components and PACES curriculum

Conversation patterns

Agents in trialogues generate questions, hints, feedback, corrections, explanations, and guidance on using the system Conversation modes: Testing, helping the peer, game competition Minimal natural language input from adult

- Multiple media
- Practical texts and tasks for adults

AutoTutor Trialogue

Teacher Agent: Cristina





Peer Agent: Jordan

To Raise the Minimum Wage in America

Economics paper written by Andrew Pink for Intro to Economics at the University of Antarctica

Do you think living on \$7.25 per hour is enough to survive in America? Do you think there should be a wage increase, say, to \$9 per hour? Well I think there should be an increase in the level of minimum wage. I think it's important to allow people to earn more money, that way they can live comfortably- live above the poverty line. Also, I read in Forbes that people who are currently under paid would

Voon their inhe if they receive an increase and this will he

What is the topic of the article?

Poverty

Government issues

Minimum wage

Participants read text and answer question

Repeat

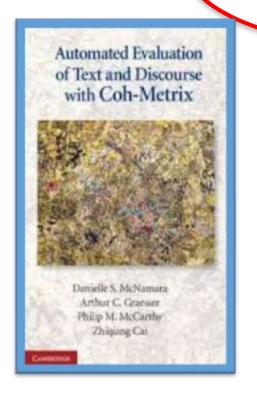


Tutorial On Digital Literacy

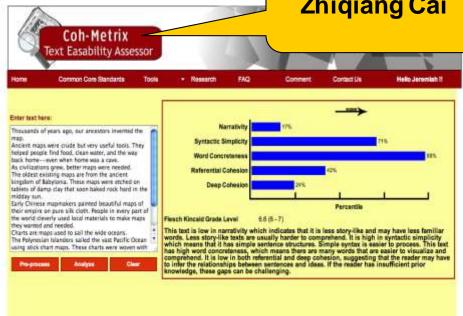


Discourse Formality

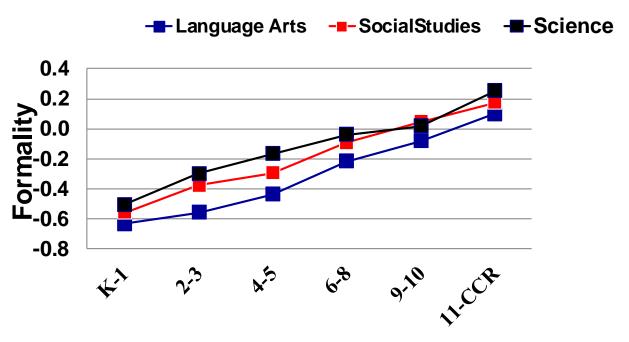
Informational Genre + High Cohesion + Complex Syntax + Abstract Words



Cohmetrix.com Zhiqiang Cai



Coh-Metrix Formality Scores as a Function of Genres and Grade Levels



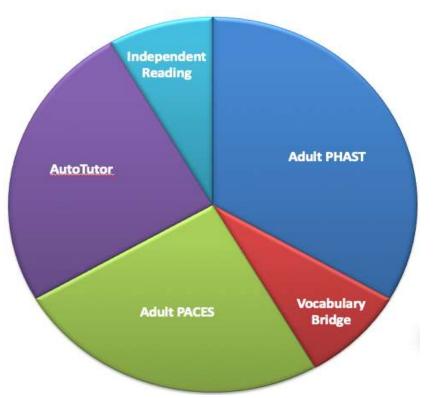
Grade Level

Graesser, A.C., McNamara, D.S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *Elementary School Journal*. *115*, 210-229.

Types of Adaptivity

- Lessons start out at a medium level of difficulty and branch to easy or hard depending on performance.
- In the inner loop (VanLehn, 2006), the conversational moves depend on the input of the adult learner.
- In the game competitions, the peer agent's actions always end up losing to the adult learner at the end.

Intervention Design



- I. Pretest on dozens of measures
- II. Intervention (100 hours)
- III. Posttest with dozens of measures

Three comprehension measures

- Woodcock-Johnson
- Sara (Educational Testing Service)
- Lexia (formerly Rapid)

Study on Adult Readers (N=253)



Overall Results of AutoTutor

Completion of lessons

- 26 lessons were used in the analysis
- 68.2% of lessons attempted
- 55.3% of the lessons completed

Performance in answering questions

- 68% answered correctly on first attempt
- 78% likelihood of branching to difficult texts rather than easy texts/items after performing on medium texts/items
- 32.8 seconds per question

Cluster Analysis

Features

 Performance time & accuracy crossed with four levels of reading comprehension (Words, Textbase, Situation model, Rhetorical structure)

Method

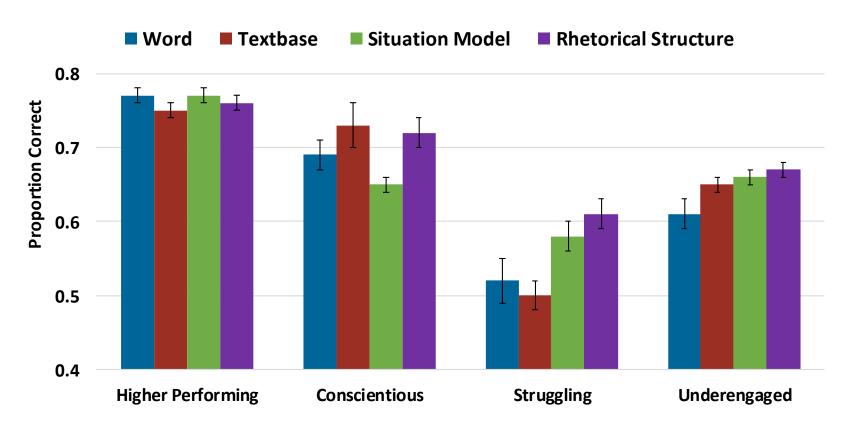
- K-means clustering VS hierarchical agglomerative clustering (Connectivity, Silhouette Width, Dunn Index)
- Hierarchical clustering performed better

Fang, Y., Shubeck, K.T., Lippert, A., Cheng, Q., Shi, G., Feng, S., Gatewood, J., Chen, S., Cai, Z., Pavlik, P. I., Frijters, J.C., Greenberg, D., Graesser, A. C. (2018). Clustering the Learning Patterns of Adults with Low Literacy Interacting with an Intelligent Tutoring System. In K.E. Boyer & M. Yudelson (Eds.), *Proceedings of the 11th International Conference on Educational Data Mining* (pp.348-354). Buffalo, NY: Educational Data Mining Society.

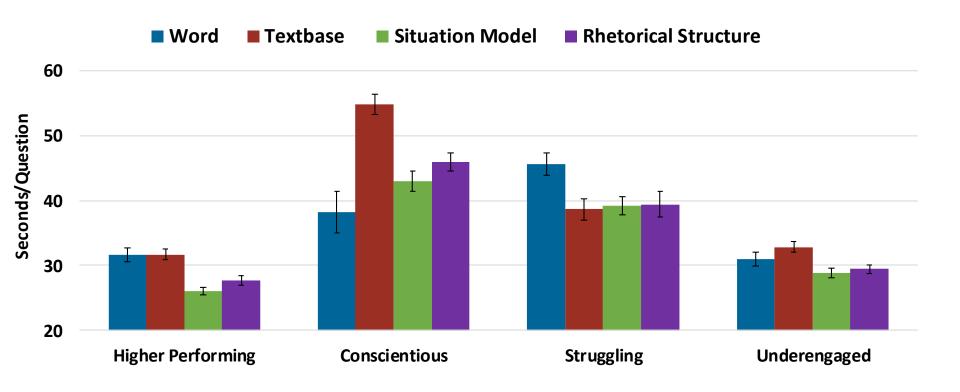
Four Clusters of Readers based on AutoTutor Response Times and Accuracy

- 1. Struggling readers showed minimal gains and may be wheel spinning. Slow plus inaccurate.
- 2. Under-engaged readers don't spend quite enough time that they need. Fast and lower accuracy.
- 3. Conscientious readers are slow and higher accuracy.
- 4. Higher performing readers are relatively fast and accurate.

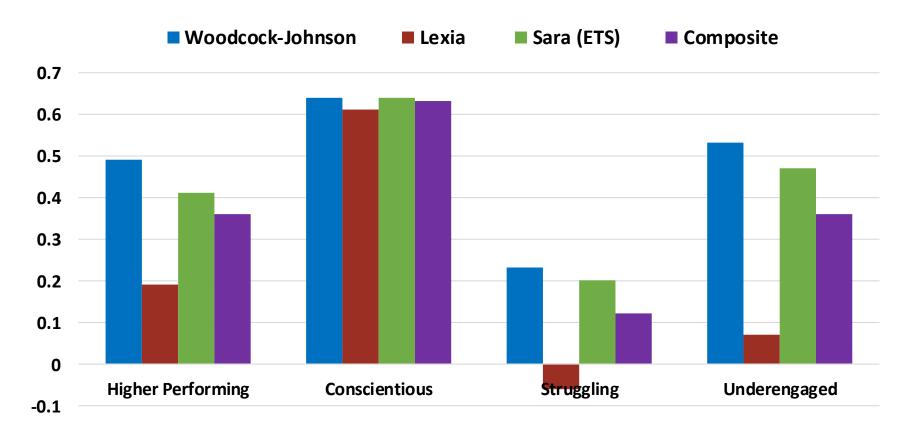
Accuracy as a Function of Theoretical Level and Reader Cluster



Time per Question (seconds) as a Function of Theoretical Level and Reader Cluster



Effect Sizes (posttest minus pretest) on Learning Gains as a Function of Reader Cluster



The four clusters of readers show very different profiles

- 1. Struggling readers showed minimal gains and may be wheel spinning. The intervention is beyond their zone of proximal development.
- 2. Under-engaged readers need to be encouraged to spend more time concentrating or otherwise be motivated more.
- 3. Conscientious readers are the major beneficiaries.
- 4. Higher performing readers may benefit from more challenge and be encouraged to increase reading activities.

Item Analyses: Mixed-effects Models (Ying Fang)

Predictor variables

- Text formality score
- Question depth level (Bloom's taxonomy)
- Answer length (log of number of words in options)

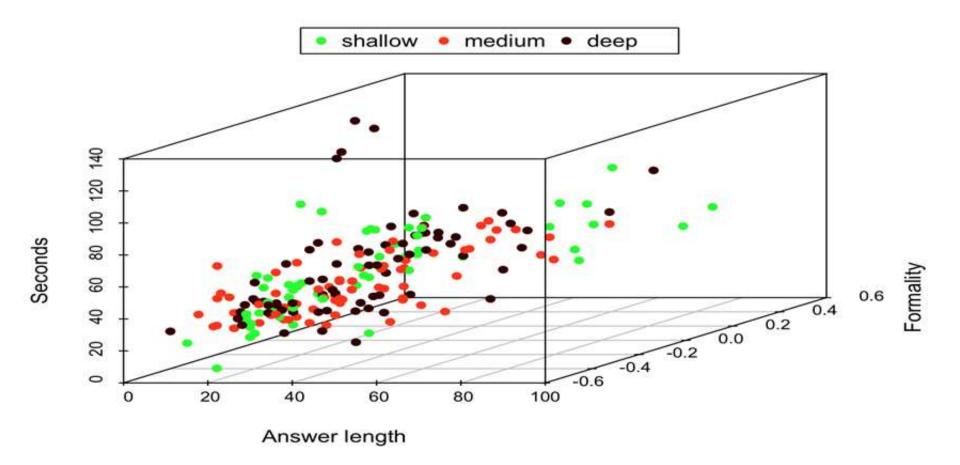
Dependent variables

- Time on question in seconds
- Correctness of answer to question

Random effects

- Participants
- Texts

Predicting time in 1st attempts

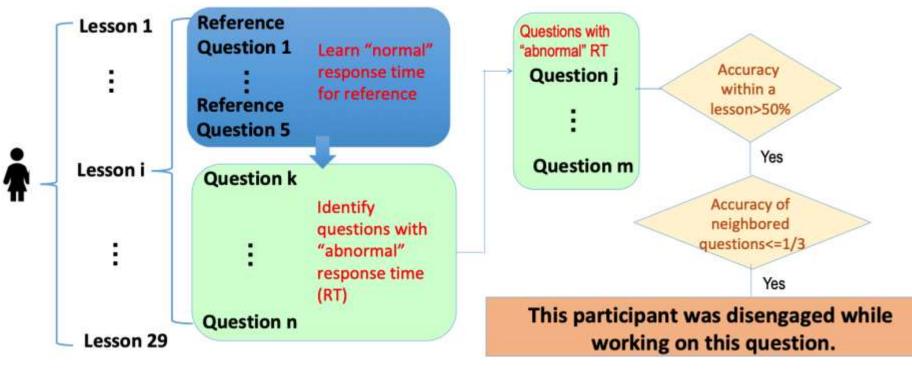


Results on Time and Accuracy of Question Items

	Text Formality	Question Depth	Answer Length
First Attempt Time	+	+	+
Subsequent Attempt Time		+	+
First Attempt Accuracy	-		-
Subsequent Attempt Accuracy		-	-

Disengagement Tracing System (Su Chen)

Algorithm Flow Chart



Upshot of Question Item Analyses

- Identify disengagement time spans of individual readers
- Quickly classify individual readers into one of the four reader clusters (struggling, under-engaged, conscientious, higher performers)
- Design AutoTutor to select materials and trialogue moves that are sensitive to these characteristics

Immediate Next Steps

- AutoTutor in the Wild
- Explore how much human instructor scaffolding is needed?
- Scaling up AutoTutor for:

adult literacy centers

workforce

colleges, universities, Department of Defense training

Long-term Horizon with Al

- Build a more adaptive AutoTutor
 - Sensitivity to engagement and reading clusters
 - Sensitivity to interests of the reader (Andrew Olney)
- Speech recognition and mobile devices
- Social connections through social media:
 - Human connections with peers and instructors
 - Will they believe intelligent bots?
- Integration with Geographical Information Systems
 - You are near a literacy center, bookstore, library...
 - You like Chinese food. What do you think about this restaurant?

Faculty
Investigators

Zhiqiang Cai, M.S, University of Memphis
Jan Frijters, Brock University
Art Graesser, University of Memphis
Xiangen Hu, University of Memphis
Hongli Li, Georgia State University
Maureen Lovett, University of Toronto
Lee Branum-Martin, Georgia State University
Robin Morris, Georgia State University
Andrew Olney, University of Memphis

Daphne Greenberg, Georgia State University

Partners

Christopher D. Barr, University of Houston Larry Condelli, American Institutes for Research Stephanie M. Cronen, American Institutes for Research

Project Managers

Whitney Baer, University of Memphis Anne Lippert, University of Memphis Dina Schwam, Georgia State University Amani Talwar, Georgia State University

Others

17 Graduate Students
22 Staff Members
25 Expert Consultants
74 Research Assistants





Institute for Intelligent Systems

THE UNIVERSITY OF MEMPHIS

Thank you!

graesser@memphis.edu