Learning by Communicating with AutoTutor, Trialogues, and Other Pedagogical Agents

Art Graesser

Professor, Psychology & Institute for Intelligent Systems, University of Memphis
Honorary Research Fellow, University of Oxford
IIS: Integrating Methodologies
Unique interplay of Theoretical Research (cognitive science/psychology), technology development (computer sciences/engineering), educational practice, and empirical evaluation
Overview

- Memphis Agent Environments
- Impact on Learning: AutoTutor
- Impact on Learning: Trialogs
Memphis Agent Environments


Computer literacy
*(AutoTutor: Graesser, Lu, et al. 2004)*

Physics
*(AutoTutor: Graesser, Chipman, Haynes, & Olney, 2005)*

English literacy
*(Greenberg, Graesser, & Lovett, 2012)*

Physics
*(DeepTutor: Rus, D'Mello, Hu, & Graesser, 2013)*

Biology
*(Guru: Olney, D'Mello, Person, Cade, Hayes, Williams, Lehman, & Graesser, 2012)*

Critical thinking
*(ARIES: Millis, Forsyth, Butler, Wallace, Graesser, & Halpern, 2011)*
When a car without headrests on the seats is struck from behind, the passengers often suffer neck injuries. Why do passengers get neck injuries in this situation?
AutoTutor Dialogue Mechanisms

AutoTutor-Style (EMT) Dialog

Dialog with student cover expectations & correct misconceptions

“Suppose a boy is in a free-falling elevator and he holds his keys motionless in front of his him and he lets go. What will happen to the keys? Explain why.”

“I bet you can explain this a bit more.”

“What about the acceleration of the objects involved?”

“Try this. What direction are the objects going? The objects are falling _________?”

“Great job!”

“Not exactly. What force is acting on both objects?”

“Let’s back up, the force of gravity on both objects is ______.”
Speech Act Hierarchy

Human-Human kappa = .80
Human-Computer kappa = .73
Functions of Conversational Agents

- Help when initiated by the user
- Navigational guide
- Modeling action, thought, and social interaction
- Adaptive intelligent conversational dialog
- Many roles: peers, tutor, mentor
Center for the Study of Adult Literacy
What is one of the specific uses of this drug?

- Give the body nutrients
- Relieve sneezing and runny nose
- Relieve headaches
AutoTutor with ALEKS Mathematics ITS
Development of **GIFT**: Generalized Intelligent Framework for Tutoring

- **Learner Model**
- **Domain-specific knowledge**
- **Instructional Strategies**
- **Experimental System**

**GIFT**
- Modify learner model
- Performance, retention, enhanced skills, etc.
- Tutor vs. Traditional Tutoring
- Intervention vs. Non-Intervention
- Comparison of learner models
- Comparison of instructional tactics
- Ablative tutoring studies

**Empirical Evaluation of Learning Outcomes**

- Optimize strategies
- Adapt content
Impact on Learning: AutoTutor


Learning Conceptual Physics
(VanLehn, Graesser et al., 2007)

Four conditions:

- Read Nothing
- Read Textbook
- AutoTutor
- Human Tutor
Multiple Choice Tests on Deep Learning (Computer Literacy and Critical Thinking)


Storey, Kopp, Wiemer, Chipman, & Graesser (2009)

---

**Computer Literacy**

- Read nothing
- Read Textbook
- AutoTutor

**Critical Thinking**

- Read nothing
- Read Textbook
- AutoTutor

Adjusted post-test scores
How does AutoTutor compare to comparison conditions on tests of deep comprehension?

- **0.80 sigma** compared to pretest, doing nothing, and reading the textbook
- **0.22** compared to reading relevant textbook segments
- **0.07** compared to reading succinct script
- **0.13** compared to AutoTutor delivering speech acts in print
- **0.08** compared to humans in computer-mediated conversation
- **-0.20** compared to AutoTutor enhanced with interactive 3D simulation

**ZONE OF PROXIMAL DEVELOPMENT**
## Mastery of a Core Concept over Time and Tasks

<table>
<thead>
<tr>
<th></th>
<th>Pretest Essay</th>
<th>Pretest MC</th>
<th>Training</th>
<th>Posttest Essay</th>
<th>Posttest MC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All-or-none Learning</strong></td>
<td>X00X</td>
<td>X0XXXX0XXX</td>
<td>X0XX0X1XX1</td>
<td>X1X1</td>
<td>XX1XX1XXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XXX0XXX0XX</td>
<td></td>
<td></td>
<td>X11XXXXXXX1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XX0X0X</td>
<td></td>
<td></td>
<td>XX1XXX</td>
</tr>
<tr>
<td><strong>Variable Learning</strong></td>
<td>X10X</td>
<td>X0XXXX0XXX</td>
<td>X0XX1X1XX0</td>
<td>X1X0</td>
<td>XX1XXXXXX1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XXX1XXX0XX</td>
<td></td>
<td></td>
<td>XX1XXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XX0X0X</td>
<td></td>
<td></td>
<td>XX1XXX</td>
</tr>
<tr>
<td><strong>No Learning</strong></td>
<td>X00X</td>
<td>X0XXXX1XXX</td>
<td>X0XX0X1XX0</td>
<td>X1X0</td>
<td>XX0XXXXXX1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XXX0XXX0XX</td>
<td></td>
<td></td>
<td>X10XXXXXXX0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XX1X0X</td>
<td></td>
<td></td>
<td>XX1XXX</td>
</tr>
<tr>
<td><strong>Refresher Learning</strong></td>
<td>X00X</td>
<td>X0XXXX1XXX</td>
<td>X1XX1X1XX1</td>
<td>X1X1</td>
<td>XX1XX1XXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XXX1XXX1XX</td>
<td></td>
<td></td>
<td>X11XXXXXXX1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>XX1X1X</td>
<td></td>
<td></td>
<td>XX1XXX</td>
</tr>
</tbody>
</table>
Emotions During Learning
(Graesser, Baker, Craig, D’Mello, Lehman, Rodrigo)

- Boredom (23%)
- Confusion (25%)
- Delight (4%)
- Flow (28%)
- Frustration (16%)
- Surprise (4%)
Emotion Sensors and Channels

- Face
- Posture
- Speech
- Dialogue
AutoTutor Comparisons (D’Mello & Graesser)

- Original AutoTutor
- Affect-sensitive AutoTutor
  - Supportive
  - Polite
- Shake-up AutoTutor
  - Rude
    (Not even close. You must be day dreaming)
  - Acknowledges student emotions
    (You’re obviously frustrated.)
  - Confrontive and face-threatening
What did we find?

• Affect-sensitivity is not the panacea for all the ills of learning
  – Benefits low domain knowledge students when they need support
  – Initially detrimental to high knowledge students

• There is a time for emoting
  – Start with regular tutor until problems become evident
  – Be supportive after 30 minutes for low domain knowledge
  – Shakeup after 1 hour for high domain knowledge (?)
Impact on Learning: Trialogues


Operation ARIES!

Scientific reasoning

Electronic textbook

Periodic questions – answer and ask

Critique experiments and newspaper clips

Game context with aliens

Trialogs with tutor and student agents
Core Science Principles & Potential Flaws

- Poor or missing comparison group
- No random assignment
- DV could be more sensitive, accurate, or precise
- DV is not scored objectively
- DV is not valid
- Subject bias
- Mortality or attrition
- Small sample size
- Poor sample selection
- Experimenter bias
- Premature generalization of results
- Confuse correlation with causation
Operation ARIES!

Learning Principles
- Adaptive
- Dialogue
- Feedback

Game Aspects
- Narrative
- Player Control
- Competition
An epic story . . .

- Love interests
- Political intrigue
- Surprises
- Revenge
- Mystery
- Double-agents
- Aliens
- Spies
- Saving the world
- Uncertainty
- "Green" theme
- Secret messages
- Twists and turns
- Defectors
- Humor
Trialogs in Learning

Low Ability → Vicarious learning

Medium Ability → Tutorial dialogue

High Ability → Teachable agent
Cadet Training

- E-Text
- 3 types MC?
- Adaptive Trialogue
Proving Ground

Cases

Identify flaws

Hint list
Example Cases

Psychology
Does using cell phones hurt driving?
Is a new cure for autism effective?

Biology
Do chemical and organic pesticides differ on food quality?
Does milk consumption increase bone density?

Chemistry
Does a new product for winter roads prevent water from freezing?
Does eating fish increase blood mercury levels?

Formats
Articles, advertisements, blogs, letters to editors
Active Duty

- Cases of suspects
- Ask questions
- Identify flaws
Core Concepts are presented across the game

-------------  Story elements  -------------

Interactive text  Case studies  Interrogation
Measures of Learning Process

- Time on Task
- Discrimination
- Generation

- E-book
- MC questions
- # Words in Trialog

- Case critiques
- Identifying Flaws (hits-false alarms)
- # of Words

- Interrogation Time
- Score-card (hits-false alarms)
- # of words
Assessments

- Multiple-choice
- Short Answer
- Case Based Reasoning
Proportional Learning Gains = .35

PLG = (posttest – pretest) / (1 – pretest)
<table>
<thead>
<tr>
<th>Chapter Title</th>
<th>Reliability, Precision, Accuracy</th>
<th>Validity</th>
<th>Objective Scoring</th>
<th>Control Groups</th>
<th>Random Assignment</th>
<th>Subject Bias</th>
<th>Representative Sample</th>
<th>Sample Size</th>
<th>Experimenter Bias</th>
<th>Causal Claims</th>
<th>Generalizability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean PLG</td>
<td>0.44</td>
<td>0.19</td>
<td>0.36</td>
<td>0.43</td>
<td>0.24</td>
<td>0.5</td>
<td>0.19</td>
<td>0.48</td>
<td>0.44</td>
<td>0.17</td>
<td>0.34</td>
</tr>
<tr>
<td>Words in Training Module</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination in Training</td>
<td>+</td>
<td>++</td>
<td>--</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Time in Training</td>
<td>--</td>
<td>++</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words in Case Study</td>
<td>-</td>
<td>++</td>
<td>--</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time in Case Study</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination in Case Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time in Interrogation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words in Interrogation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrimination in Interrogation</td>
<td></td>
<td>++</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

++ (significant positive correlation with PLG (p < .05))
+ (positive correlation with PLG (p < .1))
-- (significant negative correlation with PLG (p < .05))
-(negative correlation with PLG (p < .1))
Operation ARA (10 hours)

Pretest → ARA → Posttest

Cadet Training Module
Proving Ground Module
Shallow Learning (Multiple Choice)

Four conditions:

- Control
- Cadet Training
- Proving Ground
- Cadet + Proving

Learning Gains (post-pre scores)
Four conditions:

- Control
- Cadet Training
- Proving Ground
- Cadet + Proving
Case-based Reasoning (Deepest)

Four conditions:
- Control
- Cadet Training
- Proving Ground
- Cadet + Proving

![Learning Gains (post-pre scores)]
Conclusions about Assessments with Agents

- Shallow knowledge is acquired through didactic training
- Deep knowledge is learned through didactic training plus case-based applications
- Particular concepts are learned by different pedagogical methods
- Interactive agents are an improvement over reading and listening to monologues
Confusion Induction Experiments
(D’Mello, Lehman, Graesser, Pekrun, 2014)

• Two Agents present contradictory information to induce confusion and

• Promote learning!
  – Contradictions are good….
  – ….when learners are successfully confused
In the fall semester, all students in one section of a statistics course were told that the textbook was optional. All students in another section of the same statistics course were told that reading the textbook was required. The same professor taught the two statistics courses and gave the same lectures to each. The professor found no difference on the final exam scores between the two classes. So the textbook does not matter. And if it doesn’t matter, why buy textbooks?

Chris: Well, I think there’s a problem because of how the participants were put into each condition.
Dr. Williams: I completely disagree. It wasn’t problematic.
Dr. Williams: Lauren, do you think there’s a problem with how the participants were put into each group? Please type problem or no problem.
Lauren: problem
## Manipulation

<table>
<thead>
<tr>
<th>TUTOR AGENT</th>
<th>STUDENT AGENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Correct Opinion</td>
<td>• Correct Opinion</td>
</tr>
<tr>
<td>• Correct Opinion</td>
<td>• Incorrect Opinion</td>
</tr>
<tr>
<td>• Incorrect Opinion</td>
<td>• Correct Opinion</td>
</tr>
<tr>
<td>• Incorrect Opinion</td>
<td>• Incorrect Opinion</td>
</tr>
</tbody>
</table>
Conclusions

• Confusion induction is successful in about half the observations
  – Self-reported affect (retrospective review of conversation and facial expressions)
  – Learner changes mind during course of conversation on whether there is a flaw

• Confusion does causally improve learning and reasoning
  – Moderating gains
  – If successful confusion induction