


The Effectiveness of Tutorial Dialog in an Automated Conversational Tutor



Kristen E. Link, Victoria Pomeroy,
Rachel DiPaolo, Sonya Rajan, Bianca
Klettke, Laura Bautista, Roger Kreuz,
Art Graesser, and The Tutoring
Research Group

The University of Memphis

Overview

- Human-to-human tutoring
- Introduction to AutoTutor
- Test of AutoTutor's effectiveness
- Conclusions
- For the future

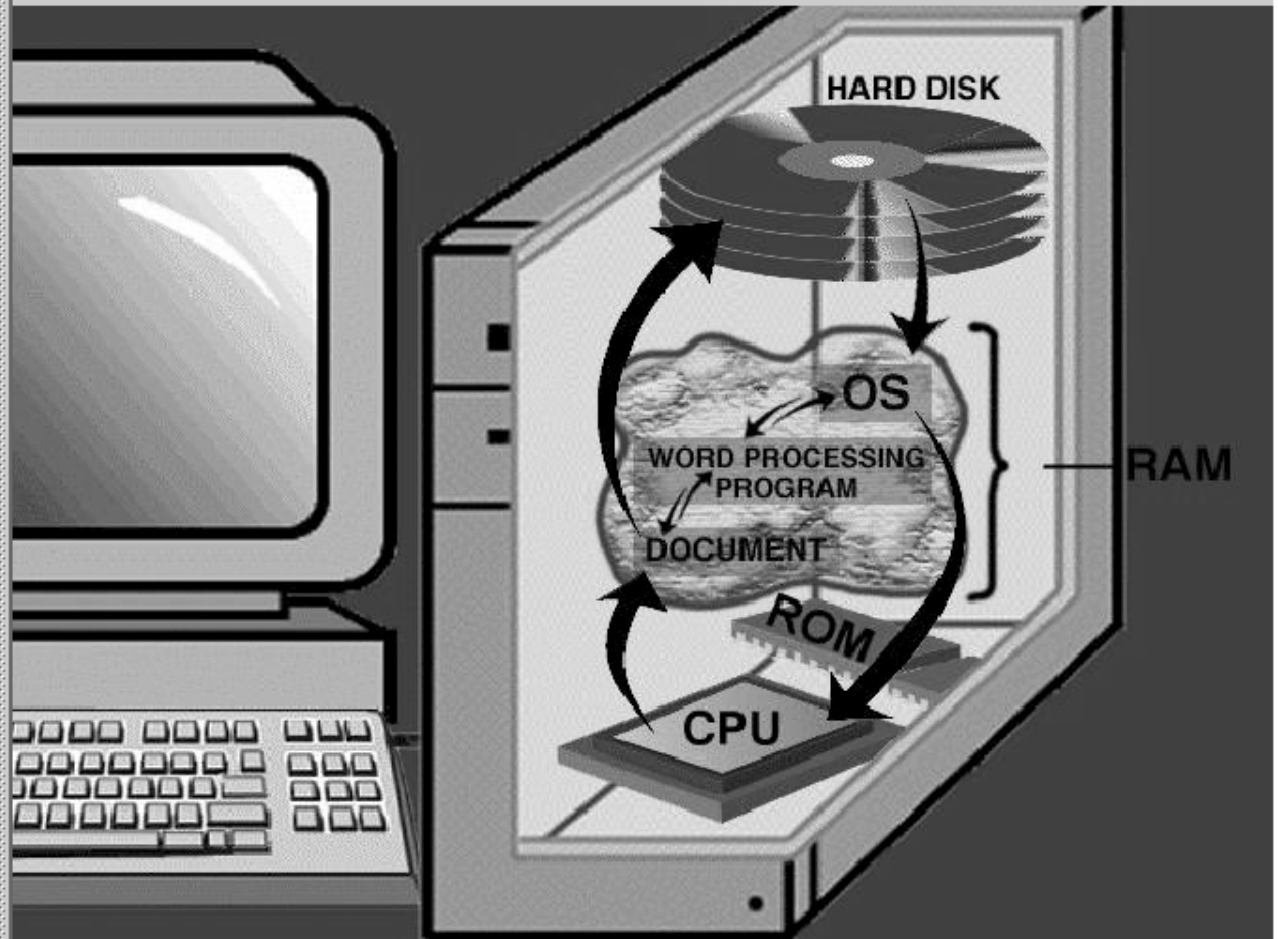
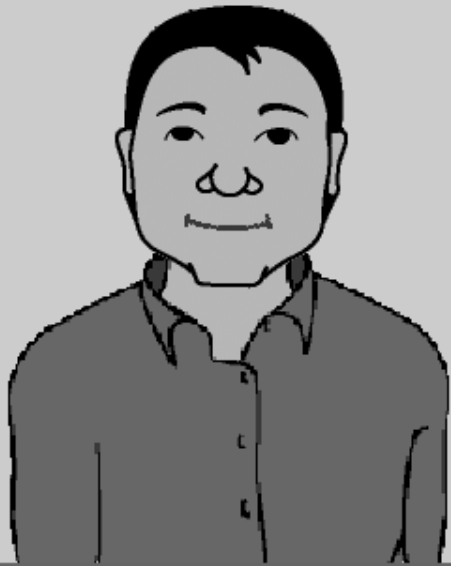
Human-to-Human Tutoring

- Pedagogically effective despite the fact that they are typically untrained (Graesser Wiemer-Hastings, Wiemer-Hastings, Kreuz, & TRG, 1999)
 - Questions/problems that promote deep reasoning
 - Collaborative, interactive discourse results in collaborative building of explanations

AutoTutor 1.0

- Simulates human tutor conversational strategies
- Synthesized speech with intonation
- Talking head with facial expressions
- Computer literacy
 - Hardware
 - Operating systems
 - The Internet

How does the operating system interact with the word processing program when you create a document?



The operating system loads the document for the application.



Modules

- Word, punctuation segmenter
- Part of speech classifier (Olde, Hoeffner, Chipman, Graesser, & TRG, 1999)
- Speech act classifier
- Latent Semantic Analysis (LSA; Landauer, Foltz, & Laham, 1998)
 - Measures quality of student contributions (e.g., match to good answers, bad answers)
- Curriculum script
 - Questions/problems, figures/diagrams, good answers, bad answers, dialog moves

Dialog Moves

- Pump (requests more information)
- Prompt (elicits specific word or phrase)
- Splice (inserts correct answer)
- Hint (presents fact or leading question)
- Elaboration (contributes important, overlooked information)
- Feedback (positive, negative, or neutral)
- Summary

Dialog Move Generator

- Type of dialog move chosen is determined by fuzzy production rules
 - Quality of student contribution (LSA)
 - Global parameters (e.g., ability, verbosity)
 - Topic coverage (each aspect of the answer)
- Examples:
 - IF good answer is high & completeness is low THEN Pump
 - IF subtopic ability is high & good answer is low THEN Hint

Evaluation of AutoTutor

- 48 undergraduates enrolled in computer literacy interacted with AT
- 3 within-subjects conditions
 - Re-read chapter(s) for one topic (e.g., computer hardware)
 - Interact with AT for a second topic (e.g., operating systems)
 - Re-read and AT for third topic (e.g., the Internet)
- 16 additional control subjects (test first)
- Survey of computer literacy

The Test

- 18 shallow multiple choice
 - The physical parts of the computer system are called:
- 12 deep multiple choice
 - How can you *best* find information on the World Wide Web?
- 18 Cloze passages (72 total)

Example Cloze Passage

The CPU, or _____, performs three main functions for the computer. The CPU runs the computer _____ by executing each program's instructions. The CPU performs the computer's mathematics. The CPU _____ with the different components of the computer, including input peripheral devices such as the keyboard and mouse, output peripheral devices such as the monitor and printer, and the computer's _____ devices such as RAM, ROM, and the hard disk.

Example Dialog

Tutor: Why is it important to consider the characteristics of the CPU [when planning to buy a computer]?

Student: speed

Tutor: And can you add to that?

Student: applications

Tutor: Okay. Well, if a program is designed for one kind of CPU, it may not run on another kind of CPU because the program and CPU are not [pause]

Student: accessible to one another

Tutor: Right, are not compatible. Okay, ...

Correlation Analyses

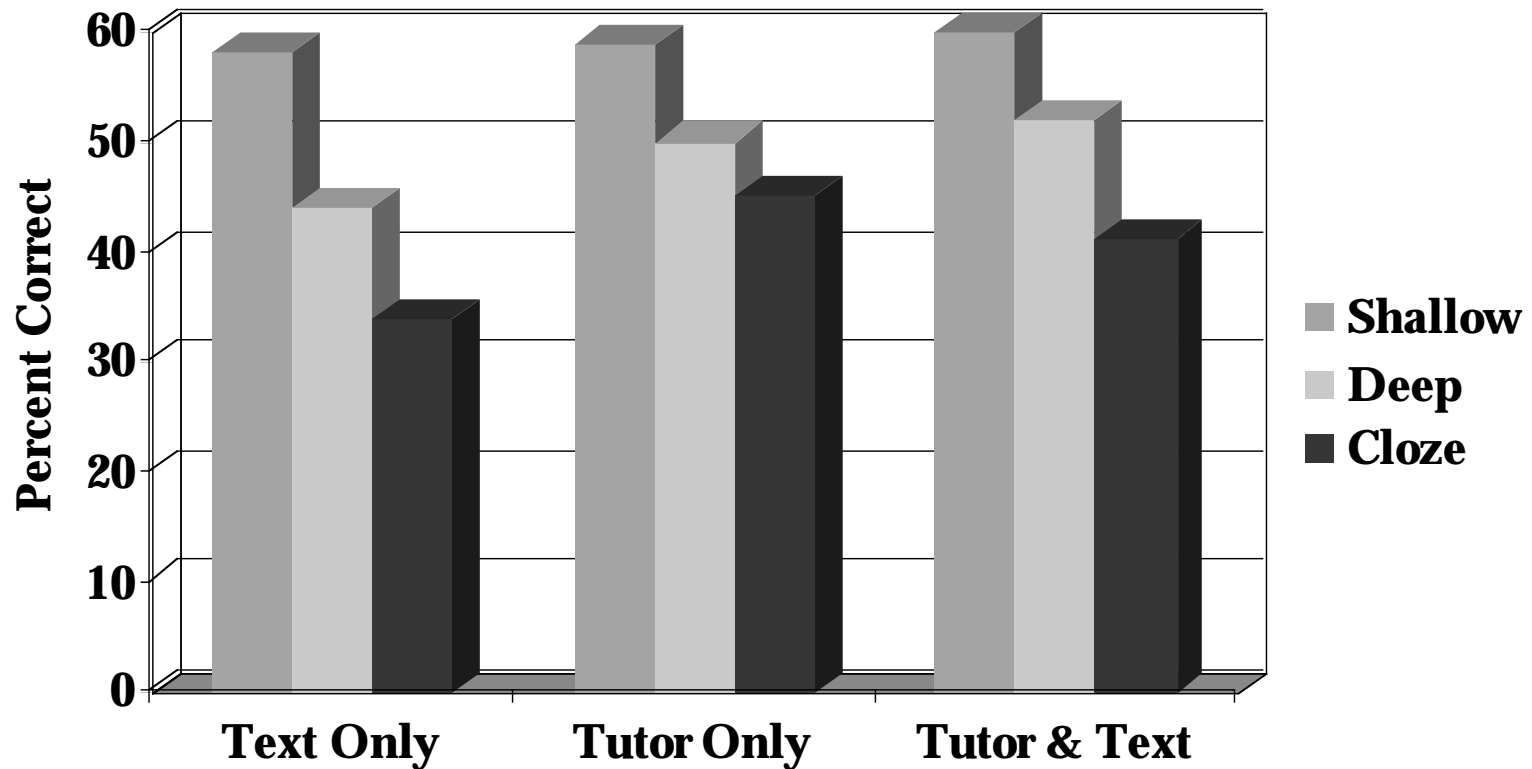
- Time spent on AutoTutor and LSA measure of curriculum script knowledge: $r(64) = .31$ ($p < .05$)
- Computer literacy score and test score: $r(58) = .32$ ($p < .05$)
- Test score and grade in class: $r(64) = .33$ ($p < .01$)

Experimental vs. Control: Performance on Test

<u>Condition</u>	<u>Mean</u>	<u>SD</u>
Experimental*	46.5%	14
Control	39.9%	10

*Subset of test questions which tested knowledge of subtopics on which participants were tutored

Mean Test Score by Condition and Question Type



Effect Size Estimates

Cohen's d (Compared to control group)

<u>Condition</u>	<u>Mean</u>	<u>SD</u>	<u>d</u>
Control	40%	10	----
Text Only	39%	15	.07
Tutor Only	47%	17	.50
Tutor & Text	46%	16	.46

Conclusions

- AutoTutor proved to be pedagogically effective
 - Effect sizes of .46 to .5
- Results not just due to practice effects
 - Tutoring helped more than re-reading
- Collaborative, conversational nature of AutoTutor the key

For the Future

- Re-read condition better controlled
 - Time spent reading
 - Read in the session
- Similar test of newer versions of AutoTutor (1.1. and 2.0)