Does Tutoring Really Have to be Intelligent?

Wendy E. Mackay
I.N.R.I.A.
Rocquencourt, B.P. 105
78153 Le Chesnay, FRANCE
wendy.mackay@inria.fr

ABSTRACT
This experiment was designed to determine whether or not tutoring is more effective if it is relevant to the user's current problems. The experimental design presented identical tutoring advice to pairs of subjects: advice was directly relevant to one subject (as determined by a human researcher monitoring a pre-specified task) and effectively random to the other.

The quantitative and qualitative results were strikingly different. On one hand, subjects learned almost all tutored commands, regardless of their relevance to their immediate activities, and rarely learned commands that had not been tutored. On the other hand, subjects were very enthusiastic about relevant tutoring and were frustrated when it seemed random: they felt it was an irritating interruption. An intelligent rule-based tutor may be unnecessary for effective learning if users can control the tutoring environment.

KEYWORDS: Intelligent tutoring, Wizard of Oz

INTRODUCTION

A intelligent tutor [1] simulates the behavior of a human tutor, who watches a student and provides appropriate advice. However, after many years of developing intelligent agents, researchers still question how much intelligence is necessary [2,3]. This experiment examines whether or not tutoring advice must be directly relevant to the student's current task, thus requiring a high level of intelligence, in order to be effective.

METHOD

The experiment used a Wizard-of-Oz technique, with a human researcher observing pairs of subjects performing a task and displaying pre-created tutoring advice on their screens. Subjects were "yoked" together, so that any advice given to one was also given to the other. The difference lay in the perceived relevance of the advice: the researcher observed only one member of the pair at a time and delivered advice based on her current needs. The other subject received the same advice at the same time, but it was perceived as random with respect to her current activity.

Three sets of commands were randomly assigned to the control and experimental conditions. For any subject, one set of commands never received any tutoring advice (the control condition). Tutoring that was always relevant to the user's current task was given for the second set of commands, while tutoring that was perceived as random or irrelevant was given for the third set of commands. The experimental design ensured that pairs of subjects received identical amounts and timing of tutoring: the sole difference was whether or not the advice was directly relevant to the subject's current behavior.

Subjects

Six female subjects had professional experience (from 1-1/2 to 4 years) with both a text editor and a word processor.

Software and Equipment

The researcher, in one room, could watch each subject (located in two different rooms) via a video monitor. We created a new text editor that differed from the subjects' familiar editors (written in TECO) as well as software that enabled the researcher to pop-up pre-created tutoring advice on both subject's screens. The researcher could also watch two monitors that showed the current state of each subject's screen. All keystrokes and the video of each subject, were recorded for later analysis.

Procedure

Training: All subjects received standardized minimal instructions on how to insert, delete and move the cursor. No other commands were presented.

Task: Subjects were given the same task, designed to be as tedious as possible to encourage learning new commands. Each of sixteen highly-repetitive activities could be accomplished with one or two commands.

Tutoring advice: The researcher could send brief tutoring messages to the top of each subject's screen. For example:

To delete one line of text, press the PF3 key.

Conditions: The experiment used a within-subjects design with a yoked control (fig. 1). Both subjects interact with the text editor while the human "wizard" watches subject 1. When the researcher identifies a situation in which advice on a particular command might be useful to subject 1, she pops that advice up on both subjects' screens. Subject 1 is considered to be in a "relevant" tutoring condition, whereas subject 2 is in the "yoked" condition.

Subjects were grouped into three pairs and participated in four one-hour sessions over four consecutive days (Table 1). Subjects did not receive any tutoring on any commands on days 1 and 4. Pairs A&B received "pre-tutoring" on days 2 and 3: they both received randomly-delivered advice on two of the three sets of commands during the first five minutes of the session, after which no further advice appeared.

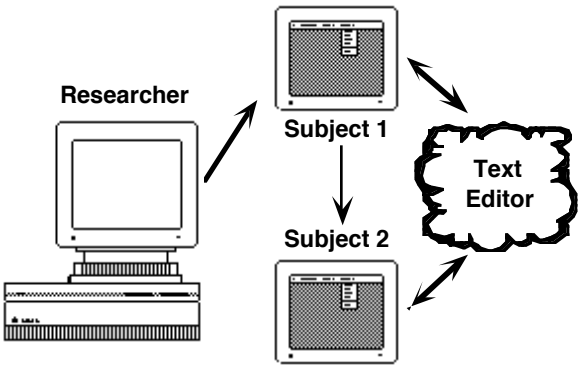

Figure 1: Wizard of Oz setup with a yoked control.

During the first 30 minutes of day 2, pair C&D received advice on delete commands, related to C's behavior. They then received advice on move commands, related to D's behavior. The order was reversed on day 3. Pair E&F was counter-balanced so that move commands were tutored first (related to E's behavior), followed by the delete commands (related to F's behavior). This was reversed on day 3.

Subject Pairs	Day 1 60 min	Day 2 30 min	Day 2 30 min	Day 3 30 min	Day 3 30 min	Day 4 60 min
A	No Tutor	Pre-Tutor	Pre-Tutor	Pre-Tutor	Pre-Tutor	No Tutor
B	No Tutor	Pre-Tutor	Pre-Tutor	Pre-Tutor	Pre-Tutor	No Tutor
C	No Tutor	Delete: Related	Move: Yoked	Move: Yoked	Delete: Related	No Tutor
D	No Tutor	Delete: Yoked	Move: Related	Move: Related	Delete: Yoked	No Tutor
E	No Tutor	Move: Related	Delete: Yoked	Delete: Yoked	Move: Related	No Tutor
F	No Tutor	Move: Yoked	Delete: Related	Delete: Related	Move: Yoked	No Tutor

Table 1: Experimental conditions

Post-Test and debriefing: At the end of day 4, the researcher interviewed each subject about the editor, the tutor and what they had learned. The researcher also met the subjects individually one week later for a follow-up interview and an informal test of their knowledge of the text editor.

RESULTS

If intelligent tutoring is necessary, then subjects should have learned more when tutoring was "intelligent", i.e. directly related to their current tasks. We found that subjects did need tutoring advice: they learned few or none of the commands in the non-tutoring control condition. In contrast, all subjects learned most or all of the commands in each of the three tutoring conditions (*pre-tutor*, *related* or *yoked*). However, there were no statistically significant differences in amount learned ($p<.01$), nor in commands remembered a week later ($p<.01$). Tutoring was clearly more effective than no tutoring, but the two forms of "random" tutoring were not statistically significantly different from the "intelligent" tutoring.

The quantitative analysis implies that there is no reason to create an expensive rule-based tutor, since randomly-delivered tutoring would be equally effective. However, the qualitative data tells a different story.

In our debriefing interviews, users did not comment on either the no-advice or pre-tutoring conditions. However they had strong emotional reactions as they experienced the change from one yoked experimental condition to the other. Although they did not realize that the "tutor" was a person, their comments were strongly anthropomorphic. One subject described a disturbing change from the *related* to the *yoked* condition, when tutoring no longer appeared relevant to her actions: "At first the computer liked me ... then it seemed to get mad at me!" Another subject, who started in the *yoked* condition, said that the computer got much smarter over the course of each session. All subjects said they enjoyed the *related* condition and were frustrated in the *yoked* condition: the tutor became an irritating interruption that they wanted to turn off.

DISCUSSION

The study suggests that providing tutoring advice can help users learn new commands, but questions whether a sophisticated rule-based tutor is required. The quantitative data, when considered alone, suggest that any form of randomly-presented tutoring will help users learn. However, the qualitative data suggest that people will only respond positively to such advice if it is perceived as relevant and not as an annoying interruption.

How can we design effective tutors? Rather than concentrating on which rules to create for analyzing what the subject is doing, it seems more profitable to investigate when and how best to interrupt people with advice. We suggest two low-cost alternatives. The first is allow users the chance to see hints about different functions at the start of an editing session (as in the *pre-tutor* condition here). Such advice must be easy to turn off and on (across work sessions) and the level of difficulty could be based on a log of the user's commonly-used commands or chosen by the user. The second approach would imitate the *yoked* condition with short bits of advice appearing randomly throughout a session. However, users must be able to easily turn this on and off: sometimes one is ready to learn new things and sometimes it is important to just get the job done.

ACKNOWLEDGMENTS

Thanks to M. Fineblum for help running the experiment and to M. Beaudouin-Lafon for his comments.

REFERENCES

[1] Anderson, J.R., Corbett, A.T., Koedinger, K.R. & Pelletier, R. (1995) Cognitive Tutors: Lessons Learned. J*ournal of the Learning Sciences*, 4, 167-207.

[2] Birnbaum, L., Horvitz, E. Kurlander, D., Lieberman, H., Marks, J. & Roth, S. (1996) Compelling Intelligent User interfaces: How Much AI? In *Proceedings of the 1997 ACM International conference on Intelligent Interfaces* (Orlando, FL).

[3] Corbett, A. & Trask, H. (2000) Instructional Interventions in Computer-Based Tutoring: Differential Impact on Learning Time and Accuracy. In *Proc. of CHI'2000*, The Hague, Amsterdam. ACM Press, pp. 97-112.