Searching for Variables and Models to Investigate Mediators of Learning from Multiple Representations

Martina A. Rau
Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA
marau@cs.cmu.edu

Richard Scheines
Department of Philosophy
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA
scheines@cmu.edu

ABSTRACT

Although learning from multiple representations has been shown to be effective in a variety of domains, little is known about the mechanisms by which it occurs. We analyzed log data on error-rate, hint-use, and time-spent obtained from two experiments with a Cognitive Tutor for fractions. The goal of the experiments was to compare learning from multiple graphical representations of fractions to learning from a single graphical representation. Finding that a simple statistical model did not fit data from either experiment, we searched over all possible mediation models consistent with background knowledge, finding several that fit the data well. We also searched over alternative measures of student error-rate, hint-use, and time-spent to see if our data were better modeled with simple monotonic or u-shaped non-monotonic relationships. We found no evidence for non-monotonicity. No matter what measures we used, time-spent was irrelevant, and hint-use was only occasionally relevant. Although the total effect of multiple representations on learning was positive, they also had a negative effect on learning, mediated by a higher error-rate. Our evidence suggests that multiple representations increase error-rate, which in turn inhibits learning. The mechanisms by which multiple representations improve learning are as yet unmodeled.

understanding mechanisms underlying successful learning is an important educational goal.

We conducted two *in vivo* experiments to investigate the benefits from learning with a version of the Fractions Tutor that uses multiple graphical representations compared to learning with a version of the Fractions Tutor that uses only a single graphical representation. In experiment 1, students worked only with a number line (in the single representation condition), or (in the multiple representations condition) with a variety of graphical representations, including circles, rectangles, and number lines. The representations were relatively static: students could interact with the representations only by entering a number into a text field. The picture updated when the student entered the correct number. In each tutor problem, students solved a fractions problem. For instance, students were asked to add two given fractions and by typing the number of shaded sections into a text field, specifying the numerator of the sum fraction. We crossed these two conditions with a second experimental factor: whether or not students received self-explanation prompts to relate the graphical representations to the symbolic notation of fractions (e.g., ½). For example, students were asked to select "adding the number of shaded sections" to the question of what action with a circle diagram corresponds to adding the numerators using fractions symbols. Results based on an analysis of pretests, immediate posttests, and delayed posttests showed that learners significantly benefited from multiple representations, provided that they were also prompted to self-explain [

Table II gives an overview of the tutor log data for each condition. While conditions did not differ with regards to error-rate, students who received self-explanation prompts requested fewer hints than students without prompts. Students in the MGR condition with prompts spent relatively more time per step than students in the other conditions, but the differences were small.

To evaluate whether the non-monotonic variables more accurately predict students' learning, we conducted step-wise regression analyses separately for error-rate, hint---

Fig. 1. Path model for experiment 1.

In addition, we predict that in experiment 1, multiple representations (mult_rep), self-explanation prompts (se), and the interaction between multiple representations and self-explanation prompts (mr*se) predict error-rate, hint-use, and time-spent. In other words, we predict that the effects of the intervention variables are entirely mediated through students' learning behaviors. Similarly, for experiment 2, we predict that the effect of multiple representations (mult_rep) predicts error-rate, hint-use, and time-

Fig. 2. The model found by GES on data from experiment 1, with parameter estimates included. This model fits the data well: $\chi^2 = 22.11$, df = 19, $p = .29$.

Fig. 3. The model found by GES on data from experiment 2, with parameter estimates included. This model also fits the data well: $\chi^2 = 6.89$, df = 10, $p = .74$.

Fig. 2 shows a model found by GES on the data from experiment 1, with path coefficient estimates included. The model fits the data well ($\chi^2 = 22.1$, df = 19, $p = .28$), and contains a number of interesting properties. For one thing, students with higher pretest scores spend much less time per problem, but none of our intervention variables had any influence on time, and the apparent effect of time spent per step during the learning phase is minimal. Multiple representations had a positive effect on learning, but only when self-explanation prompts were also part of the learning environment.[9] Further, there is no evidence that the positive effect of multiple representations is mediated by either error-rate, hint-use, or time-spent. When not combined with multiple representations, self-explanation prompts appear to slightly increase error-rate and thus inhibit learning, but slightly decrease hint-use, which, because they appear to inhibit learning, have an overall positive effect on learning.

Fig. 3 shows a model found by GES for experiment 2 that fits the data very well ($\chi^2 = 6.89$, df = 10, $p = .74$). This model indicates that although multiple representations (mr) have a positive direct effect on both the immediate posttest and the delayed posttest, they also have a negative indirect effect on both outcomes through error-rate. Learning with multiple representations seems to cause students to make slightly more errors during learning, possibly because the greater variability in tutor problems leads to higher cognitive pro[(a n(i)6(ve)4ve)4( )f 6.44c1 Tcmakendo m,c 0 Tw 1.892

improve on the apparent monotonicity of the raw measures because our sample did not include high prior knowledge students. However, students' pretest scores covered a broad range from very low to very high (see Tables I and III). Although surprising, our findings can be taken as encouraging for the community of educational data mining and for the community of researchers who study ITSs. Analyzing raw measures of error-rate, hint-use, time-spent and learning is much easier than analyzing non-monotonic variants. Furthermore, most research that uses log data obtained from ITSs assumes monotonicity. Our findings do nothing to undermine this practice.

Our findings from path analysis modeling demonstrate the importance of model search. None of our initial hypotheses fit the data, but there are millions of plausible alternatives, only a small handful of which could be practically investigated by hand. Further, estimating path parameters with a model that does not fit the data is scientifically unreliable. Parameter estimates, and the statistical inferences we make about them with standard errors etc., are all conditional on the model specified being true everywhere except the particular parameter under test.

Even if our initial hypotheses had fit the data well, however, it would have been important to know whether there were alternatives that explained the same data. The GES algorithm implemented in Tetrad IV enabled us to find plausible models that fit the data well. The models we found in Fig. 2 and Fig. 3 allow us to estimate and test path parameters free from the worry that the model within which the parameters are estimated is almost surely mis-specified, as is the case for the model in Fig. 1.

Several caveats need to be emphasized, however, lest we give the false impression that we think we have "proved" the causal relationships that appear in the path diagrams shown in Fig. 1 and Fig. 2. First, the GES algorithm assumes that there are no unmeasured confounders (hidden common causes), an assumption that is almost certainly false in this and in almost any social scientific case, but one that is routinely employed in most observational studies.[10] In future work we will apply algorithms (e.g., FCI) that do not make this assumption, and see whether our conclusions are robust against this assumption. Second, although we did include intervention interaction in our model search for experiment 1, and did test for interactions between pretest and mediators in experiment 2, by no means were our tests exhaustive, and by no means can we rely on the assumption that the true relations between the variables we modeled are linear, as the search algorithms assume. Nevertheless, many of the bivariate relationships in the data we modeled appear approximately linear, so the assumption is by no means unreasonable. Third, we have a sample of 290 students, and although that is sizable compared to many ITS studies, model search reliability goes up with sample size but down with model complexity and number of variables, and is

[10] Although our data are from a study in which we intervened on intervention, we did not directly intervene on our mediator or outcome variables. Thus these parts of our model are subject to the same assumptions as a non-experimental study.

demonstrate, that the impact of interactive representations is an interesting question to address in future research.

In conclusion, our results are of interest both to the educational psychology literature and to the intelligent tutoring systems literature. First, we can gain insights into the effects of instructional interventions: although multiple representations seem to overall be beneficial, they also seem to lead students to make more errors during the learning phase, which is associated with lower performance on posttests. Second, once we gain knowledge about which learning behaviors are adaptive and which are not, we can use these insights to improve our tutoring systems. For example, perhaps multi-representational ITSs should be designed to prevent errors in the practice and learning phase. Perhaps we can help students avoid practice errors by providing more worked examples, or by designing better error feedback messages. Or perhaps the increase in errors is simply a cost associated with multiple representations that instructors have to live with. These questions and others arose from path analysis and model search and lead almost directly to new hypotheses that we, and hopefully others, will address in future research.

6. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, REESE-21851-1-1121307, and by the Pittsburgh Science of Learning Center which is funded by the National Science Foundation, award number SBE-0354420. In addition, we would like to thank Vincent Aleven, Nikol Rummel, Ken Koedinger, the Datashop team, and the students and teachers who participated in our study.

lvLnL(S65(t)1(hL)3(l)1, El)1.l(S65chL)3wo3(ar)3mli-1(s)5chL&oWrl
. .Hi.SLe-e-g(a)34(ndy)6Hi. D.irl(nt-)2(.)-3r(a)4(c-)3(t-)2(i)-2vr .rg(Ea)-2(via)-2riomCe(t-)2st.R(e)34viaeof(Ea)-2duc.t-irR(e)34sea stt-c-t-ir, 6r,183 -