Tis better to construct than to receive? The effects of diagram tools on causal reasoning.

Matthew W. Easterday, Vincent Aleven, Richard Scheines
Human-Computer Interaction Department, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh PA, 15213-3891

Abstract Previous research on the use of diagrams for argumentation instruction has highlighted, but not conclusively demonstrated, their potential benefits. We examine the relative benefits of using diagrams and diagramming tools to teach causal reasoning about public policy. Sixty-three Carnegie Mellon University students were asked to analyze short policy texts using either: 1) text only, 2) text and a pre-made, correct diagram representing the causal claims in the text, or 3) text and a diagramming tool with which to construct their own causal diagram. After a pretest and training, we tested student *performance* on a new policy text and found that students given a correct diagram (condition 2 above) significantly outperformed the other groups. Finally, we compared *learning* by testing students on a third policy problem in which we removed all diagram or tool aids and found that students who constructed their own diagrams (condition 3 above) *learned* the most. We describe these results and interpret them in a way that foreshadows work we now plan for a cognitive-tutor on causal diagram construction.

1. Introduction

To become effective citizens, students must be able to analyze public policy. For example, students should be able to reason about *causal questions* such as: “will decreasing the amount of junk-food commercials on TV decrease childhood obesity?” Furthermore, students’ reasoning must account for different claims presented by different sources, e.g. “researchers claim that watching junk food commercials causes obesity, while industry advocates argue that junk food commercials only affect the brand of junk food consumed, not the total number of calories.”

Taken together, several bodies of research suggest that while students have difficulty reasoning about causal arguments,[1] diagrammatic representations[2] and tools[3] might improve their reasoning. However, as Ainsworth (in press) notes,

> ... research on the benefits of providing learners with more than one representation has produced mixed results. For example, a number of studies have found that learners benefit from either constructing or being presented with [diagrams]... Unfortunately, just as many studies have shown that learners can fail to benefit from these proposed advantages of [diagrams]...

[1] See Scheines, Easterday, & Danks (in press), on causal reasoning, Kuhn (2005, 1991) on argument, and Voss, Perkins, & Segal (1991) for an overview informal reasoning in education.

[2] See Ainsworth (in press), for an overview, Harrell (2004), on argument maps, Pinkwart et. al. (2006) on legal argument diagrams, Mayer & Moreno (2002), and Bauer & Johnson-Laird (1993).

[3] See Kirishner, Shum, & Carr (2003) and Van den Braak et al. (2006), for an overviews of visualization tools and educational uses, as well as Suthers & Hundhausen (2003), and Harrell (2005).

Furthermore, the efficacy of a given diagram format interacts heavily with both the particular task on which the diagram is applied, and students' familiarity with the diagram format, i.e. the fact that a concept map can improve students' recall (Nesbit & Adesope, 2006) does not necessarily imply that a causal diagram will improve the accuracy of students' policy inferences. Thus, the following are important open questions:

- For what domains and with what kind of pedagogy do we think diagrams will help? For example, should we use diagrams to teach causal reasoning about public policy texts? Given student fluency with text in general, we may not be able to design pedagogy with diagrams that significantly improve reasoning in comparison with text, without exorbitant cost. Ainsworth (in press) shows the difficulty of learning a new diagram format in general, and only a few studies (such as Suthers & Hundhausen, 2003) have examined diagrams for causal reasoning, usually focusing on science, not policy.
- Should we give students pre-made diagrams, or should they construct their own? Some argue that students come to a *deeper understanding* of a problem or task constructing their own representations of it, while others argue that diagrams constructed by students contain too many errors to be useful, or that the empirical evidence for the benefit of construction is scant (Van den Braak, et. al., 2006). Many of the studies on diagram construction either combine construction with feedback and/or scaffolding, or do not directly contrast diagram use with diagram construction (see Cox, 1996, Grossen & Carnine 1990, and Van Meter, 2001 in Ainsworth, in press; Hall, Bailey, & Tillman 1997).
- Does using/constructing a diagram have the same effect on learning as it does on performance? Even if we can temporarily increase performance by using diagrams, students may not learn much in transfer tasks in which the correct diagram is not immediately available.

2. Task and Intervention

To explore these questions, we gave students short, fictional, policy texts as in Figure 1.

Childhood obesity is now a major national health epidemic. A number of facts are widely agreed upon by the public and scientific community: exercise decreases obesity, and eating junk food increases obesity. It's also clear that people who watch more TV are exposed to more junk food commercials.

Parents for Healthy Schools (PHS), an advocacy group which fought successfully to remove vending machines from Northern Californian schools, claims that junk-food commercials on children's television programming have a definite effect on the amount of junk food children eat. In a recent press conference, Susan Watters, the president of PHS stated that "...if the food companies aren't willing to act responsibly, then the parents need to fight to get junk food advertising off the air."

A prominent Washington lobbyist Samuel Berman, who runs the Center for Consumer Choice (CCC), a nonprofit advocacy group financed by the food and restaurant industries, argues that junk food commercials only "influence the brand of food consumers choose and do not not affect the amount of food consumed." While Mr. Berman acknowledges that watching more TV may cause people to see more junk food commercials, he remains strongly opposed to any governmental regulation of food product advertising.

Recent studies by scientists at the National Health Institute have shown that watching more TV does cause people to exercise less.

Figure 1. Policy text on obesity.

Our tasks involved answering questions like: "According to the PHS, will making kids exercise more reduce the number of junk food commercials they watch?" Note that neither *junk food commercials* nor *exercise* affects the other, so the correct answer is "no."

What is the correct diagram for a policy text and why? In the last two decades, researchers have produced a rigorous theory of causal reasoning that rests primarily on a semantics for causal diagrams in the form of directed graphs (Spirtes, Glymour, & Scheines 2000; Pearl 2000). Because the ability to use these diagrams has implications for reasoning in general and for policy reasoning in particular, teaching the theory has become somewhat of a priority. But even after weeks of instruction, students in causal reasoning courses, like those in other formal domains like algebra or physics, still fall short of being able to build and utilize formal representations reliably and accurately. This led us to wonder whether the cognitive cost of teaching causal diagrams outweighs the presumptive benefit of using or building causal diagrams from texts like Figure 1.

To test differences in performance and learning between students who used diagrams, diagramming tools, or neither (only text representations), we randomly assigned students who had no prior training in causal reasoning to one of three conditions:

1. **Text** students (the control group) received all case studies as text only (Figure 1).
2. **Diagram** students received, in addition to a text version, a correct, diagrammatic representation of the case study (Figure 2).

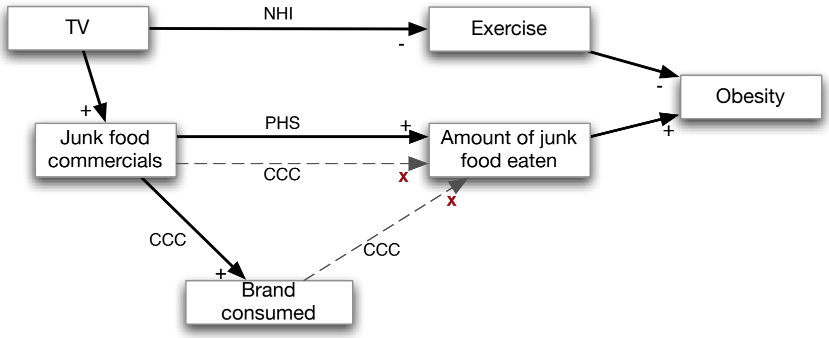

Figure 2. A causal diagram representing the case-study on obesity. Boxes represent causal variables, and arrows represent either positive (+), negative (-), or no (x) influence of one variable on another. An annotation on the arrow (e.g. PHS) identifies the source making the causal claim.

Note that to solve the question about exercise reducing the amount of junk food eaten using the diagram in Figure 2, students only need to notice that there is no set of arrows leading from the *exercise* variable to the *amount of junk food eaten* variable.

3. **Tool** students received the case study along with a computer tool with which they could construct their own diagrams (Figure 3).

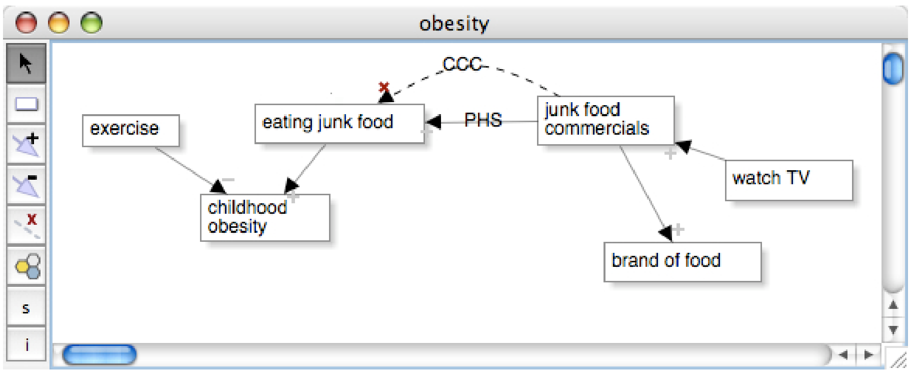

Figure 3. The iLogos tool shown in *causal* mode (Easterday, Kanarek, & Harrell, in press).

4. Participants and Setting

We investigated these questions with 63 Carnegie Mellon University students enrolled in undergraduate philosophy classes but who had no prior training in causal reasoning.

5. Research design

The study consisted of a brief (approximately 15 minute) training and three tests in which students were given a policy text and asked to answer 10 causal reasoning questions, (see Figure 4.) All students began the experiment with a policy argument on the environment (**pretest**) presented in text only. After the pretest, text students received training on causal reasoning without diagrams, while diagram and tool students received training with diagrams. We then tested *performance* (**performance test**) by giving students another policy argument on obesity presented as text to the text students, presented as text with a diagram to the diagram students and presented as text with the diagramming tool to tool students. Finally, to test *learning*, all students were given a third text on crime in which policy arguments were presented as text only (**learning test**).[4]

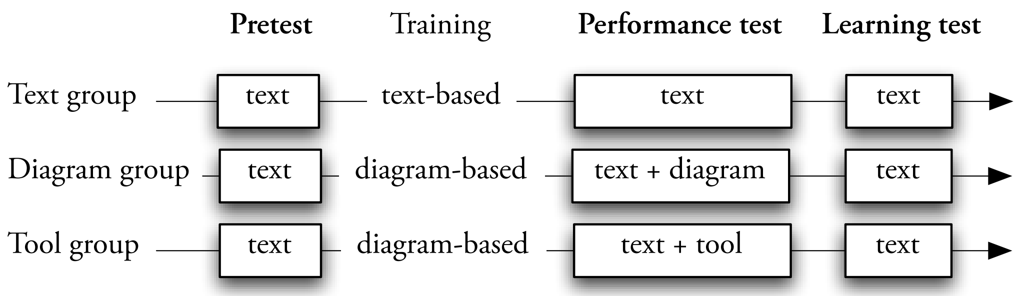

Figure 4. The experimental procedure showing order of tests and training for each group.

In the training, both groups received 4 interactive web-pages of instruction on causal reasoning. The diagram and tool groups' instruction included diagrams, while the text group's did not; to make the training as close to identical as possible, every diagrammatic explanation in the diagram/tool training was matched by an equivalent prose explanation in the text training. Diagram and tool students also received an additional page of instruction describing how the buttons of the tool worked, but with no additional information about diagrams or reasoning. While students received feedback on the problems they solved on the training, we gave them no feedback about their answers on the tests.

6. Data collection

On each test, students were asked 10 multiple choice, causal questions (e.g. "According to the PHS, will making kids exercise more reduce the number of junk food commercials they watch?"). Students could answer one of three ways, either that: a) there would be a causal effect (e.g. reducing junk food commercials *would* affect obesity), b) there would be no causal effect, or c) there was not enough information to decide. We also recorded the time students spent on each test, and whether or not they constructed a diagram on scratch paper.

[4] Note that the order of policy texts in the tests was not counter-balanced, i.e. all students received the policy text on environment in the pretest, followed by obesity on the performance test, followed by crime on the learning test. The underlying causal structure of each policy text however, (i.e. the number of variables and connections between variables), was identical across policy texts–texts differed only in cover story.

7. Results

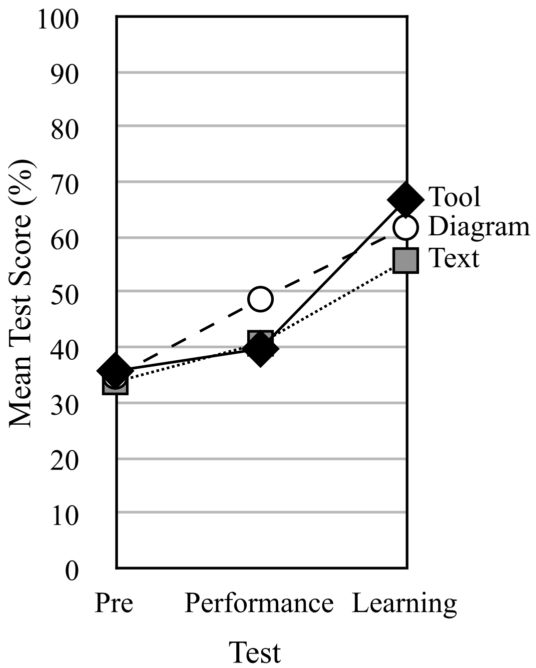

Figure 5. Mean test scores for text (n = 24), diagram (n = 24) and tool (n = 15) students on pre-, performance and learning tests.

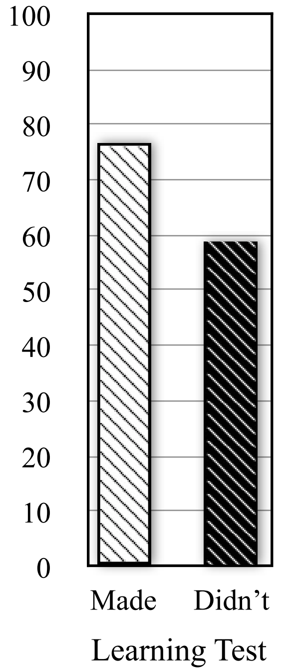

Figure 6. Mean test scores on the learning test for students who made (n = 6), or didn't make (n = 57) diagrams.

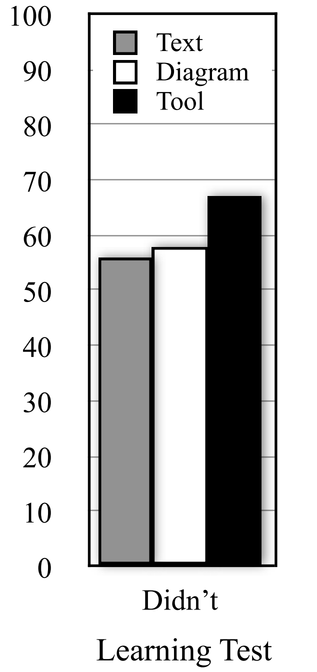

Figure 7. Mean test scores on the learning test for the students who didn't make diagrams (n = 57).

Overall performance and learning results. As can be seen in Figure 5, all groups performed at chance on the **pretest** (Text = 34%, Diagram = 35%, and Tool = 36% with no significant differences between groups). After training, students were given a performance test in which policy information was presented as text to text students, as text with a correct diagram to diagram students, and as text with a diagramming tool to tool students. On this **performance test**, diagram students scored 49%, outperforming both the text students who scored 41% ($p < .05$, $df = 54$) and the tool students who scored 40% (although the difference was not significant) showing that students reason better when text is accompanied by a correct diagram, rather than by a tool or alone.[5] On the **learning test** however, in which all students received policy information as text only, *both* tool students (67%) and diagram students (62%) outperformed text students (56%, $p < .05$, $p = .07$ respectively, $df = 57$), so while having a correct diagram seems to provide a superior advantage for improving *performance*, the dramatic gains of the tool group on the learning test undercut any clear advantage of having a diagram for *learning*.

Effect of making a diagram on the learning test. To better understand the tool group's learning gains, we looked separately at students who made diagrams and students who didn't make diagrams on the learning test (see Figure 6). The 6 students who made diagrams performed better (77%) than the 57 students who did not make diagrams (59%, $p <$

[5] We performed a multiple regression with the mean test score as the outcome variable, 2 dummy variables representing experimental condition, and time on test and training as covariates. On the learning test, we used time on training as a covariate. The multiple regression analysis is equivalent to an ANOVA and was used because the coefficients are easier to interpret.

.05, df = 61) demonstrating either the usefulness of diagrams or a selection effect showing that "good" students make diagrams.

Effect of tool practice on the learning test. The scores of students who did *not* make diagrams (Figure 7) is even more interesting; among the students who did not make diagrams on the learning test (13 tool students, 20 diagram students, and 24 text students), students in the tool condition who had practiced making diagrams on the performance test scored 67%, outperforming the students in the text condition who scored 56% ($p < .05$, df = 53), and students in the diagram condition who scored 58% ($p < .10$, df = 53). Although these results are correlational they suggest that, *when diagrams are unavailable,* having practiced making diagrams with a tool leads to an increase in *learning.*

Time. There were no significant differences in time between groups on the pretest, performance test or learning test, however, the diagram group spent longer (16.1 min) on the shared training than the text group (13.4 min, $p < .05$, df = 60). One could argue that diagrams simply make students spend longer on training, however, that explanation cannot account for the results of the learning test or for the fact that the training did not have such an effect on the tool group, whose time on training (14.8 min) was not significantly different than that of the text group. In fact, given the dramatically greater amount of practice students have had *prior* to the experiment with text as compared to diagrams, it is surprising that differences in training time were not greater. Of far greater import is the relatively short time that *all* groups spent on training and testing–whereas professors and graduate students took approximately 1-1.5 hours to complete the experiment during pilot testing, participants took an average of only 32 minutes!

Main results. In summary, there are two interesting results that require explanation: 1) diagram students showed better *performance* than text and tool students, and 2) tool students demonstrated greater learning.

8. Interpretation

How do we explain performance of the diagram students and the learning of the tool students? Using a diagram to make an inference requires at least three basic steps: 1. *comprehension*, 2. *construction,* and 3. *interpretation* (see Figure 8).

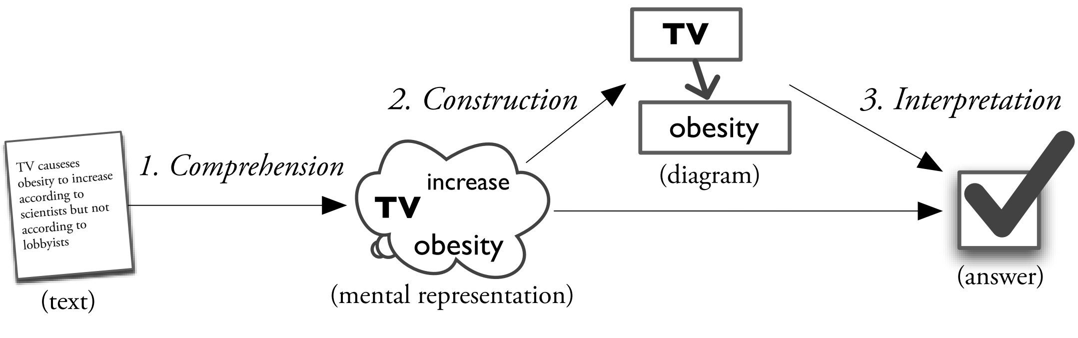

Figure 8. General steps in using an external representation.

In the *comprehension* step, (Koedinger & Nathan, 2004), students recognize a relevant piece of information in the source text, (e.g. a causal claim), presumably forming some corresponding mental representation. From that mental representation, students may either try solve the inference problem directly, (which may impose a significant working memory burden), or they may externalize that information in a new form by *constructing* a diagram. Students then *interpret* the diagram, making the inference needed to solve the problem.

If the diagram format *works*, (i.e. it reduces the required amount of working memory or number of cognitive operations), and students have learned how to use the format, then using a diagram should be more effective than using text. Also diagrams should help more than tools, because when students use a tool they must *comprehend, construct* and *interpret,* whereas students using a diagram only have to *interpret.* The process in Figure 8 thus predicts the superior results of the diagram students on the *performance* test–students had greater success with diagrams than with text or tools.

The process in Figure 8 also explains the *learning* results of the tool students: if practicing constructing diagrams improves one's *comprehension* skills, then tool students should perform better than text and diagram students when diagrams aren't used, which is exactly what we see in Figure 7. Note, that even with improved comprehension skills, solving the problem from the mental representation should still be more difficult than solving the problem with a diagram (if the diagram format has been properly designed to reduce working memory burden or the number of cognitive operations), which are the results we see in Figure 6.

To summarize, when we tested reasoning *performance*, we saw that the diagram students outperformed the others, suggesting that, indeed, the *interpretation* step (of reading off the diagram to get the answer) is easier than the combined *comprehension*, *construction* and *interpretation* steps required when one has a tool, (a conclusion supported by the observation that no tool student was able to construct a correct diagram on the performance test). On the other hand, when we tested *learning* by removing all diagrams and tools, students who had practiced constructing diagrams with tools outperformed the others, suggesting that practice constructing diagrams may improve students' comprehension skills.

9. Conclusion

We found that after only 15 minutes of training, students were able to make almost 10-20% more accurate inferences about the effects of different social policies when they had a correct diagrammatic representation,[6] and that practice constructing diagrams also improved students' reasoning by approximately 10%.[7]

So it seems that diagrams, whether constructed or received, are indeed useful for (learning) causal reasoning about public policy. With effective instruction over a period of weeks, we are hopeful that students can not only handle the sorts of problems we examined here, but also policy problems of more realistic complexity. In pursuit of this goal, we are now conducting protocol studies of diagram construction on experts and novices. We hope to leverage these studies into a cognitive model of diagram construction, and leverage the cognitive model of diagram construction into a cognitive tutor.

---

[6] Comparing the average score of 49% for diagram students to the 40% and 41% of the text and tool students on the performance test, and comparing the average score of 77% for students that made any diagram on learning test to the 59% of students who didn't make diagrams.

[7] Comparing the averages of 67% of tool students who did not make diagrams on the learning test to the 58% of diagram students and 56% of text students.

10. References

Ainsworth, S. E. (in press). DeFT: A conceptual framework for learning with multiple representations. *Learning and Instruction.*

Bauer, M. I., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science,* 4(6), 372-8.

Cox R. (1996). *Analytical reasoning with multiple external representations*. Unpublished doctoral dissertation. England: University of Edinburgh.

Easterday, M. W., Kanarek, J. S., & Harrell, M. (in press). Design Requirements of Argument Mapping Software for Teaching Deliberation. In T. Davies, & B. S. Noveck (Eds.), *Online Deliberation: Design, Research, and Practice*. Stanford, CA: CSLI Publications/University of Chicago Press.

Grossen, B., & Carnine, D. (1990). Diagramming a logic strategy - Effects on difficult problem types and transfer. *Learning Disability Quarterly,* 13(3), 168-82.

Harrell, M. (2004). *The improvement of critical thinking skills in What Philosophy Is* (Tech. Rep. No. CMU-PHIL-158). Pittsburgh, PA: Carnegie Mellon University, Department of Philosophy.

Harrell, M. (2005). Using argument diagramming software in the classroom. *Teaching Philosophy,* 28(2), 163-77.

Hall, C., Bailey, J., & Tillman, C. (1997). Can Student-Generated Illustrations Be Worth Ten Thousand Words? *Journal of Educational Psychology,* 89(4), 677-81.

Kirschner, P. A., Buckingham Shum, S. J., & Carr, C. S. (Eds.). (2003). *Visualizing argumentation: Software tools for collaborative and educational sense-making*. London: Springer-Verlag.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences,* 13(2), 129-64.

Kuhn, D. (1991). *The skills of argument*. Cambridge, MA: Cambridge University Press.

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

Larkin, & Simon (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65-99.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, 12, 107-19.

Nesbit, C., & Adesope, O. (2006). Learning with concept and knowledge maps: A meta-analysis. Review of Educational Research, 76(3), 413-48.

Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge Univ. Press.

Pinkwart, N., Aleven, V., Ashley, K., & Lynch, C. (2006). Toward legal argument instruction with graph grammars and collaborative filtering techniques. *ITS 2006, LNCS 4053*, 227-36.

Scheines, R., Easterday, M., & Danks, D. (in press). Teaching the Normative Theory of Causal Reasoning. In A. Gopnik, & L. Schultz (Eds.), *Causal learning: Psychology, philosophy, and computation*. Oxford, England: Oxford University Press.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search* (2nd ed.). Cambridge, MA: MIT Press.

Suthers D, Hundhausen C. (2003). An Empirical Study of the Effects of Representational Guidance on Collaborative Learning. *Journal of the Learning Sciences,* 12(2), 183-219.

Van den Braak SW, Van Oostendorp H, Prakken H, Vreeswijk G. (2006, August). A critical review of argument visualization tools: do users become better reasoners? Paper presented at the ECAI-06 Workshop on computational models of natural argument, Trento, Italy.

Van Gelder, T. J. (2003). Enhancing deliberation through computer supported visualization. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.), *Visualizing argumentation: Software tools for collaborative and educational sense-making* (pp. 97-115). London: Springer-Verlag.

Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology*, 93(1), 129-40.

Voss, J. F., Perkins, D. N., & Segal, J. W. (1991). *Informal reasoning and education*. Hillsdale, NJ: Lawrence Erlbaum.