When observation beats doing: Learning by Teaching

Sandra Y. Okita, Daniel L. Schwartz
School of Education, Stanford University
485 Lasuen Mall, Stanford, CA 94305-3096
yuudra@stanford.edu; danls@stanford.edu

Abstract: Forty adult participants tested the hypothesis that an important aspect of learning-by-teaching is the opportunity to watch one's student perform. Participants studied a passage on the body's mechanisms for causing fever. They then completed one of four conditions. (a) Teach and then observe their student answer questions. (b) Teach and then self-study the same questions oneself. (c) Self-study and then observe a student answer questions. (d) Self-study and then self-study again. Results indicated that teaching and observing one's student led to greatest learning gains both for the questions one's student tried to answer and new questions that had not been raised. In some cases, it is better to observe than do.

Introduction

"Learning by Doing" is often an optimal way to help students learn (Barron et al., 1998). Underneath its broad umbrella, one can find variants that range from constructionism (Kafai & Harel, 1991), to problem-solving (Anzai & Simon, 1979), to project-based learning (Krajcik, 2001a; Krajcik et al., 2001b). Learning by doing can mean many different things, but for our purposes, the key contrast is between "learning by doing something oneself" versus "learning by observing somebody else do the same thing."

A priori, there are many benefits that should accrue from learning by doing oneself. People remember better when they generate instead of receive (Slamecka & Graf, 1996). People also develop a better sense of a problem space when they have open-ended exploration (Vollmeyer, Burns, & Holyoak, 1978). Feedback is directly coupled to one's actions, plus students can learn meta-skills associated with self-regulated learning. There are also important motivational benefits. Students are more inclined to persevere and find satisfaction when they are engaged in doing. In contrast, passively observing somebody else do something is often considered a sub-optimal way to learn (though sometimes unavoidable, for example, when a class requires a lecture).

Despite the enthusiasm for learning by doing, there may be situations where learning by observing is superior. Social psychology, for example, includes many cases where people learn by watching and imitating other people's behaviors (Bandura, 1977; Bandura et al., 1973 Meltzoff, 2002). Thus, people can learn by observing other people. The question we pose is whether there are situations where observing another person is actually a more effective way to learn than doing the same thing oneself? And if so, what are the types of learning content and conditions that make this possible.

There are many examples of observational learning in the implicit learning literature, such that people do not intentionally try to learn and they are largely unaware of any learning that may be taking place. Implicit learning has been found to be effective in domains including attitudes, motor skills, and language acquisition. Regardless, observational learning through implicit mechanisms is probably less relevant for the kinds of topics most educators need to teach. Many educators need to teach conceptual topics that require students to explicitly learn relations among relatively abstract (and often complex) ideas. Thus, we confine our question to conceptual topics that are explicitly included in many curricula. (This is not to say that implicit learning is not critically important in educational settings.) It also seems non-contentious that observation can be better than doing, when students cannot accomplish a task, but the person they are observing can. The person can provide a model of competent performance. In the study below, we demonstrate something more surprising. We show that students can learn better by observing, when the person they observe knows less than they do. Thus, our study is more about learning among peers than learning from expert models.

How could observing somebody who knows less (or knows about the same) be beneficial for students? At a minimum, we assume that observers need to have sufficient "pre-understanding" to make sense of what they observe. We also assume that students need a predisposition to observe closely; it is not difficult to imagine students day dreaming and missing a potential learning moment altogether. However, if students have a good pre-understanding and predisposition, they might spontaneously compare their understanding to what they observe in another person, and any discrepancies can alert them to think more deeply about who is right. Under the right circumstances, observing another person might be a great trigger for reflection.

Learning by teaching may create ideal circumstances for learning by observing. As people teach their pupil, they develop a "pre-understanding" of the domain they are teaching, plus they develop some understanding of what their pupil knows. When they subsequently observe their pupil answer questions or perform, they can readily compare their ideas and expectations with what they observe. Plus, they may have high motivation to see how their pupil performs. As an analogy, a coach may be quite engaged and learn a great deal when observing her players perform in a game.

Learning by teaching typically involves three phases. We present them sequentially, though in reality, they can be interwoven. The first phase involves preparation for teaching. Research has indicated that students who prepare to teach others can learn more than students who prepare to take a test themselves (Bargh & Schul, 1980; Biswas et al., 2001). The second phase involves actually teaching. If we assume interactive teaching (not a stand-up lecture), there is growing evidence that the process of teaching and explaining can be an effective way to learn (Chi, 2001; Graesser, Person, & Magliano, 1995; Palinscar & Brown, 1984). Webb (1989), for example, found that during group work, students who provide explanations learn more than the students who receive them. In this case, it would appear that doing (the explanation) is better than observing (the explanation). However, there is a third phase of teaching that we think may be particularly good for learning by observation. This third phase occurs when teachers observe whether their pupils can effectively use what they were taught to solve problems or complete tasks. If teachers do not have perfect knowledge of the domain, their pupils' performance can reveal gaps in what the teacher taught and perhaps understands. The pupil serves as projective feedback. Plus, the performance of the pupil may provide alternatives the teacher did not think of. Even if these alternatives are not correct, they may slow down the teacher's natural inertia to keep thinking in the same way.

To the best of our knowledge, the literature on learning-by-teaching and small group interaction has not separated out the "final" observational phase of learning-by-teaching from the broader act of teaching. Thus, it is unknown whether there is any special value to observing one's pupil, and certainly there is no direct evidence comparing whether this observation is better than doing it oneself. There is, however, promising indirect evidence from research on Teachable Agents (Schwartz et al., in press). Teachable Agents are a form of pedagogical computer agent that we created on the premise that students can learn a great deal by teaching. Students teach their agent. Based on what it has been taught, the agent can then answer questions using simple artificial intelligence techniques. Students can observe their agent's answers and depending on the quality of the answer, students can revise the agent's understanding (and their own). In one study, we examined the value of having the student observe their agent's performance. We compared a hundred high school students learning hypothetico-deductive reasoning using a Teachable Agent. In the Do group, students (a) induced various rules, (b) deduced and tested predictions from those rules, and (c) created a standardized representation of each rule. In the Teach group, students (a) induced various rules, (b) created a standardized representation of each rule, and (c) watched their agent use the representation to deduce and make predictions. Thus, students in both conditions completed the same activities with the exception that students in the Teach condition observed their agent using the rules, and the students in the Do condition used the rules themselves. A week later, students completed a paper and pencil test of hypothetico-deductive reasoning. Students in the Teach condition significantly outperformed the students in the Do condition.

These results are suggestive, because they neatly separate the process of doing and observing, and they showed that observing was more effective than just doing. However, the observation phase involved looking at a computer program, not another person. Therefore, we thought it was important to conduct a second study with human pupils. The study was designed to answer the following question: Would participants learn more by observing their students answer questions compared to how much they learned when answering the same questions themselves? If the answer is yes, then the results would show that observing can be more valuable than doing.

There is an inherent difficulty in implementing this comparison experimentally. As the participant observes the student (whom we will call Student X), Student X may introduce important information that provides an inadvertent advantage to observation over doing something oneself. To control for this possibility, we constructed a relatively elaborate experimental design and procedural setup. We set it up so that some students never taught, but they still observed Student X answer questions. The Student X they saw was the exact same Student X that the "teaching" participants saw. Participants watched a canned videotape of Student X answering questions. In other words, the exact same information was available to those participants who observed Student X (their student) and those who observed Student X (some student). If Student X does not provide any special information, then participants who observe Student X (but never taught her) should look about the same as participants who never observe Student X.

In all, there were four conditions. Figure 1 provides a schematic of the design and procedural flow. Our prediction was that teaching and then observing Student X (one's own student) would lead to better learning and understanding compared to (a) teaching but then doing things oneself; (b) doing things oneself (but not teaching) and then observing Student X; and (c) doing things oneself and then doing things oneself again. To help separate the effects of the conditions, we had focal questions at each session that we could assess during a posttest. For example, while some participants were teaching Student X around a fixed set of questions, other participants were studying those questions themselves. The posttest also included questions that did not appear during the various stages of teaching and observing. These questions were highly inferential, and therefore, they assessed overall understanding rather than picking up specific facts that participants learned when teaching and observing.

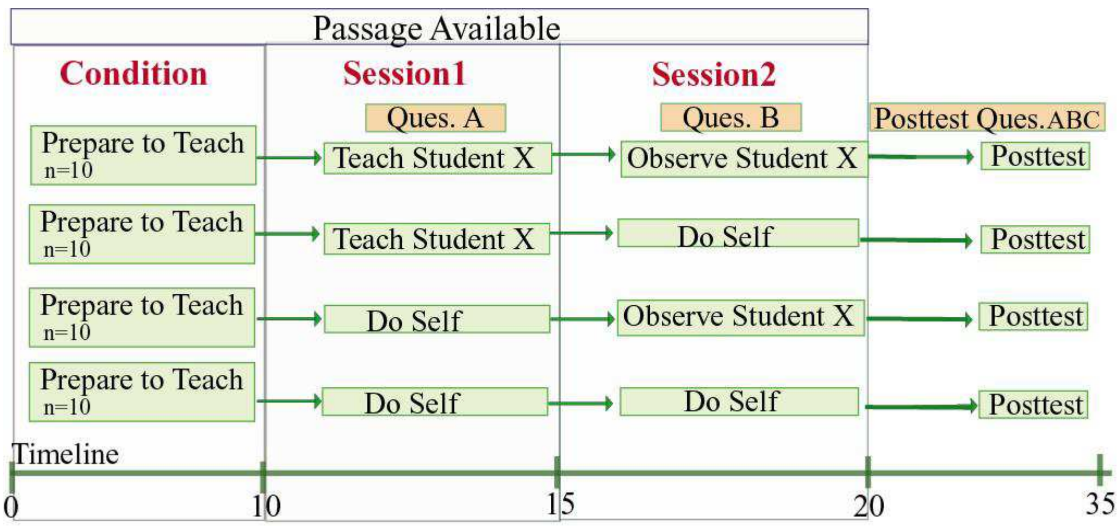

Figure 1. Procedural Flow and Design.

Method

Participants

Forty (20 female, 20 male) graduate students voluntarily participated in the study. They were randomly assigned to one of four conditions.

Design and Procedure

The study comprised four conditions created by crossing the two factors of Teach and Observe (see Fig. 1); Teach+Observe, Do-Self+Observe, Teach+Do-Self, Do-Self+Do-Self. For the Teach factor, which occurred first, participants either taught Student X or worked on their own (Teach v. Do-Self). For the Observe factor, which occurred second, participants either observed Student X answer some questions, or they answered the questions themselves (Observe v. Do-Self). During the Teach/Do-Self session participants worked with Question Set A (described below). During the Observe/Do-Self session, students worked with Question Set B. Participants never received feedback on any answers given to the Questions Sets. Afterwards, participants completed a post test by answering Question Set A, Question Set B, and a new Question Set C.

Participants completed the study individually. Participants studied a one-page passage on fever for 10 minutes. They were told they should prepare to tutor a student about fever. Participants had the passage at their disposal throughout the study, except at posttest. After preparing, participants entered their experimental condition. In the Teach condition, participants met Student X. Student X was a confederate who was used for all participants, and who studiously avoided introducing new information and targeted questions. The participants also received

Question Set A. They were told that their student needed to be tutored around these questions along with any other facts the participant felt the student should learn. In the Do-Self condition, participants received Question Set A and were told to study them with the passage. The participants tutored or studied Question set A for about 5 minutes. After completing the Teach/Do-Self session, participants entered the Observe/Do-Self session which also lasted for 5 minutes. In the Do-Self condition, participants received Question Set B and were asked to study them along with the passage. In the Observe condition, participants watched a videotape of Student X trying to answer Question Set B. The questions were asked by another confederate playing the role of examiner. Student X did not provide any complete answers and the examiner did not provide feedback. Before seeing the videotape the participants heard that Student X might be right or wrong, so they "might want to look along with the passage." Participants who had taught Student X thought they were seeing their student answer the questions. Participants who had not taught thought they were seeing a student answer the questions. After completing the treatments, the participants completed a posttest where they orally answered Question Sets A, B, a new Question Set C.

Materials and Measures

The fever passage explained how the human body gets and maintains a fever. It explained the mechanisms that trigger the fever response (e.g., macrophages), the mechanisms that introduce more heat into the body (e.g., shivering), and the mechanisms that prevent the body from releasing heat (e.g. blocking sweat).

Each Question Set had five questions. Question Sets A and B were largely factual (e.g. "What processes cause the body to increase temperature?"). Question Set C used inference questions (e.g. "Why does a dry nose mean a dog might have a fever?"). We scored each question on 0 to 2 point scale (1: incorrect/no answer, 2: partially correct but incomplete, 3: precise and detailed). Thus, for any Question Set, the maximum score is 10. Table 1 provides a sample scoring.

Table 1: Scoring Method

<table>
<tr><th colspan="3">Scoring Method (0-2 point scale)</th></tr>
<tr><td>0: incorrect/no answer</td><td>1: partially correct but incomplete</td><td>2: precise and detailed</td></tr>
<tr><td colspan="3">Why is shivering not enough to create a fever?</td></tr>
<tr><td>0 point:</td><td colspan="2">"Because its not enough, you need more"</td></tr>
<tr><td>1 point:</td><td colspan="2">"Because shivering alone creates heat, but the brain is not involved so it doesn't set the temperature set point."</td></tr>
<tr><td>2 points:</td><td colspan="2">"You can create heat with shivering, but you also need a mechanism that doesn't let that heat escape, so you need the hypothalamus to raise the set point."</td></tr>
</table>

Results

We had three leading hypotheses. The first was that students who taught would do better than students who did not teach on Question Set A (which occurs in the Teaching v. Do-Self session). This hypothesis received moderate support. The second, more central hypothesis was that students who taught and observed their student would do better than the other three conditions on Question Set B. Set B is the content of the Observation/Do-Self session. This hypothesis received good support. The third hypothesis was that the students who taught and observed would also do better on Question Set C. Set C involves inferential questions that only appeared on the posttest and evaluate overall depth of understanding. This hypothesis also received good support. Evidently, observing one's student answer questions is quite effective in supporting learning, even more so than addressing the exact same questions oneself or observing a "non-pupil" answer those questions. Figure 2 shows the average scores for participants in each condition on each question set.

To test these effects statistically, we conducted a multivariate analysis (MANOVA). Teach (Teach/Do-Self) and Observe (Observe/Do-Self) were crossed between-subjects factors. Question Set (A, B, C) was a within-subject factor on the dependent measure of test accuracy. An omnibus test indicated a significant three-way interaction of Teach X Observe X Test; $F\ (2, 35) = 3.3$, Hotelling's T, .19, $p < .05$. Planned contrasts indicate that this effect derives from the relative gain of the Teach+Observe condition on Question Sets B & C compared to the other conditions; $F(1, 36) = 4.9$, $p < .05$. When aggregating all three Questions Sets into a single measure, there is also a main effect of the opportunity to Teach compared to Do-Self, as one would expect from the literature on learning-by-teaching; $F(1, 36) = 16.6$, $p < .01$. Looking at Figure 2, one can see that both Teach conditions show relatively superior performance across the three problem sets There is also an overall main effect of Observe compared to Do-Self; $F(1,36)=7.29$, $p<.05$. However, this latter result is largely due to the Teach+Observe condition and cannot be taken as a strong endorsement of the idea that observing works in all cases.

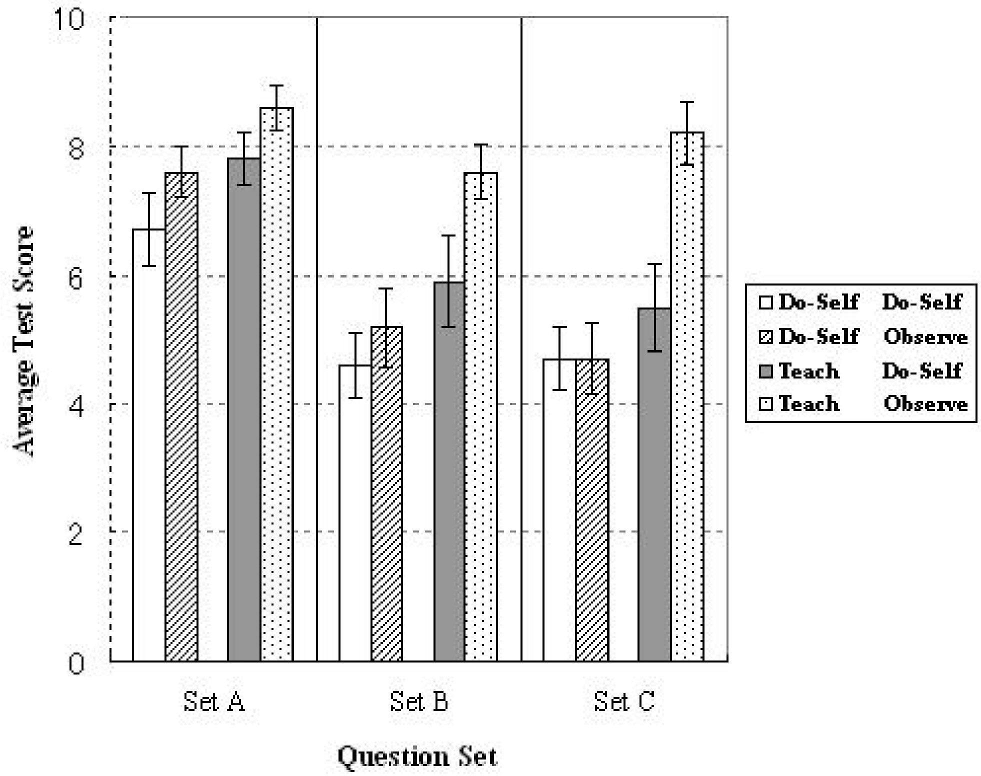

Figure 2. Average Score by Condition

One way to make sense of these results is to compare the Teach+Observe condition with the Do-Self+Observe condition. In both conditions, the participants watched the same student try to solve Question Set B, and therefore, they were exposed to the same information. Similarly, neither group of participants had any experience with Question Set C, which only appeared on the posttest. Participants who observed their own student averaged 79% of the total points across the two sets of questions compared to 49% for the students who observed a student they had not taught. It is also informative to note that students who did Question Sets A & B by themselves (Do-Self+Do-Self) were descriptively the worst.

Discussion

The current experiment isolated one element of learning-by-teaching. This element is the chance to see how one's student performs. We hypothesized that this type of observation can be highly valuable for learning, even more so than doing the same tasks oneself. The results were amazingly in accord with a set of detailed predictions (at least we are amazed given how complex the predictions were and how unlikely it is for the results to fall out just so).

First, we found that teaching other students around a set of questions led to gains compared to just working on those problems oneself. These results showed up, even though the confederate student did not contribute very much information in the teaching interaction besides statements like, "I think I understand", and "Oh, I didn't know that." Notably, the benefits of teaching lasted into the later portions of the study, so that even the teaching participants who subsequently studied on their own (instead of observing) showed advantages on questions from the later parts of the study. Thus, the results support the basic idea that interacting with a pupil can be helpful for learning.

Second, we found that teaching and then observing one's pupil has powerful effects on what one learns. For Question Set A the Teach+Observe condition showed a moderate advantage, which makes sense because these are the questions from the teaching session. The Teach+Observe condition started to truly separate itself on Question Set B. Question Set B was covered when these participants observed their student performing, and thus, it is exactly where we would have predicted the differences to start showing up. Interestingly, the benefits carried over to the questions in Set C. This suggests that the Teach+Observe participants had developed a fuller model of temperature regulation. Importantly, the students in the Do-Self+Observe condition observed the same video of the student working on Question Set B. However, they showed minimal benefits for observing compared to the otherwise comparable students who worked on Question Set B themselves (the Do-Self+Do-Self condition). Evidently, the participants who had taught their student had been prepared to learn by observing their student subsequently.

These novel findings naturally raise a number of important questions for subsequent research. In addition to questions of replicability, there are questions about generality. For example, would it work with another population, another topic, or a less controlled situation? Another class of questions asks about the mechanisms behind these effects. While we are not in a position to address mechanism questions based on this study, one might entertain two classes of mechanisms. One is cognitive and one is affective. We believe both are necessary and that each includes numerous processes.

One cognitive account might go something like the following: Self-monitoring is a complex cognitive task. For example, in the context of problem solving, it requires one "process" that applies a sequence of steps to solve the problem, and a second "process" that evaluates the problem solving process. For instance, when doing a math problem, one ideally runs a systematic process that computes a precise answer, and a second process that does a quick and dirty analysis to estimate an approximate answer. In essence, people need to run and coordinate two processes that work on solving the problem; the one doing the "actual work" and the one that evaluates the work to see what changes may need to be made to improve the results. If the answers are too far out of alignment, this should trigger a set of activities to resolve the discrepancy. By hypothesis, the dual task demands of self-monitoring can be alleviated by monitoring somebody else's work. One does not need to run two processes, but instead, can apply the monitoring process to somebody else's thinking. This would free up capacity and permit more effective reflection and debugging. A second, but not incompatible, cognitive account might be that observing another person provides new issues and questions that help overcome the inertia of one's own ideas. Azmitia (1996) mentions, for example, that peers can help one revise knowledge by providing information that one would have otherwise never considered or by requiring explanations from another point of view, there are other possibilities too.

A limitation of a purely cognitive account is that the participants who observed someone they had not taught did not learn as well as students who observed their own pupil. This suggests something of an affective account, because the "freeing up of capacity" story should apply even if Student X is not one's pupil. By the affective account, the participants who observed their own student answer questions were more engaged and locked into their pupil's performance. They felt some responsibility or relation towards the student and therefore paid more attention to her answers. A hybrid cognitive-affective account is also possible. The participants had knowledge of what they had taught the student. Several participants, for example, spontaneously showed dismay when they saw their student get a question they had forgotten to teach her about. When they heard the questions the student received, they realized they had not taught the relevant ideas, which may have caused them to think it through more deeply. Because they had taught, Student X was their mirror.

One way to get at the affective versus cognitive accounts would be to run a study where all the students teach a student. Afterwards half of them get to see their student answer questions and the other half get to watch a student they did not teach (even though both students would give the same answers). If an affective connection is a critical ingredient, then one would expect the participants who see their own student to learn more.

Ultimately, we are not in a position to pinpoint the causes of the "observation advantage." Our suspicion is that there are not one or two causes. We believe the effect is due to a felicitous confluence of many important mechanisms that favor learning. These would include a sense of connection, responsibility, reflection, over-the-shoulder monitoring, attention, concurrent modeling, and so on. This is one of the reasons that learning-by-teaching can be such a powerful instructional method. It is a single situation that naturally brings to bear very many positive forces for learning.

References

Anzai, Y., Simon, H. A. (1979). The Theory of Learning by Doing, *Psychological Review*, 86(2), 124-140.

Azmitia, M. (1996). Peer interactive minds: developmental, theoretical, and methodological issues. In Baltes, P. B. and Staudinger, U. M. (Eds.), *Interactive Minds: Life-span perspectives on the social foundation of cognition* (pp. 133-162). Cambridge University Press**.**

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., & Zech, L. et al. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *Journal of the Learning Sciences,* 7(3-4), 271-311.

Bandura, A. (1977). *Social Learning Theory*. New Jersey: Prentice-Hall, Inc.

Bandura, A., Jeffery, R.W. (1973). Role of symbolic coding and rehearsal processes in observational learning, *Journal of Personality and Social Psychology*, 26, 122-130.

Bargh, J., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology,* 72, 593-604.

Biswas, G., Schwartz, D.L., Bransford, J.D., & TAG-V. (2001). Technology support for complex problem solving: From SAD environments to AI. In K. Forbus & P. Feltovich (Eds.), *Smart machines in education* (pp.71-98). Menlo Park, CA: AAAI/MIT Press.

Chi, M.T.H., Silver, S.A., Jeong, H., Yamauchi, T., & Hausmann, R.G. (2001). Learning from human tutoring. *Cognitive Science*, 25, 471-533.

Devin-Sheehan, L., Feldman, R. S. (1976). Research on Children Tutoring Children: A Critical Review. *Review of Educational Research*, 46(3), pp. 355-385.

Graesser, A.C.,Person, N., & Magliano, J. (1995). Collaborative dialog patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychologist*, 9, 359-387.

Kafai, Y., Harel. I. (1991). Learning through design and teaching: Exploring social and collaborative aspects of constructionism. In I. Harel & S. Papert (Eds.), *Constructionism*. Norwood, NJ: Ablex, 1999.

Krajcik, J. (2001a). Supporting Science Learning in Context: Project Based Learning. In Tinker, R., & Krajcik, J.S. (Eds.), *Portable Technologies: Science Learning in Context*, Netherlands: Kluwer Publishers.

Krajcik, J. & Starr, M. (2001b). Learning Science Content in a Project-based Environment In Tinker, R., & Krajcik, J.S. (Eds), *Portable Technologies: Science Learning in Context*, Netherlands: Kluwer Publishers

Meltzoff, A.N. (2002). Elements of a developmental theory of imitation. In A.N. Meltzoff & W. Prinz (Eds.), *The Imitative Mind: Development, Evolution, and Brain Bases* (pp. 19-41). Cambridge, UK. Cambridge University Press.

Palincsar, A.S., Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension monitoring activities. *Cognition and Instruction*, 1, 117-175.

Schwartz, D. L., Blair, K. P., Biswas, G., Leelawong, K., & Davis, J. (in press). Animations of thought: Interactivity in the teachable agents paradigm. To appear in R. Lowe & W. Schnotz (Eds). *Learning with Animation: Research and Implications for Design*. UK: Cambridge University Press.

Slamecka, N.J., Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 592-604.

Vollmeyer, R., Burns, B.D., & Holyoak, K.J. (1996). The impact of goal specificity on strategy use and the acquisition of category structure The impact of goal specificity on strategy use and the acquisition of category structure, *Cognitive Science*, 20: 75-100 .

Webb, N.M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, 13, 21-39.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. REC-0231946, & SLC-0354453. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.