The Effects of Quality Self-Explanations on Robust Understanding of the Control of Variables Strategy

Michael A. Sao Pedro, Janice D. Gobert & Patrick Sebuwufu
{mikesp, jgobert, sebuwufu}@wpi.edu

Department of Learning Science and Technology
Worcester Polytechnic Institute
100 Institute Rd. Worcester, MA 01609

Abstract: In previous studies we compared the effectiveness of three pedagogical approaches on the acquisition and robust understanding of the control of variables strategy (CVS). In our first study, Sao Pedro, Gobert, Heffernan & Beck (2009), we showed that at immediate posttest, two *direct* learning conditions, with and without explanation, significantly outperformed a *discovery* learning condition in remediating multiply confounded experiments. In our second study, Sao Pedro, Gobert & Raziuddin (2010), we retested students six months later and found that performance at this skill dropped for those in the direct-no explain condition. However, the direct+explain condition students maintained their performance. Given the efficacy of the direct+explain condition, we investigated in this work which instructional components of this condition predicted acquisition, transfer, and long-term retention of CVS. We found that the quality of self-explanations generated while critiquing setups during practice tasks was the most predictive of robust CVS understanding, above and beyond other typical predictors such as pretest score and time-on-task.

Introduction

Within the recent debate on the efficacy of discovery learning versus direct instruction (cf. Kuhn, 2005; Kirschner, Sweller, & Clark, 2006; Hmelo-Silver, Duncan, & Chinn, 2007), some researchers question whether one pedagogical approach yields more successful acquisition and transfer of knowledge and skills than the other. For example, Klahr and Nigam (2004) compared pedagogical approaches for teaching the control of variables strategy (CVS), an important scientific inquiry sub-skill (NSES, 1996). CVS encompasses both the procedural knowledge of how to design a controlled experiment, and the conceptual understanding that designing such experiments enables the making of valid inferences about the effects of one independent variable on an outcome (Chen & Klahr, 1999; Kuhn, 2005). In Klahr and Nigam's study, students in a direct instruction condition significantly outperformed those engaged in discovery learning on a near-transfer test of CVS. These results suggest that the purported benefits of discovery learning, particularly the deeper learning, may not always occur. Dean Jr. and Kuhn (2006), however, challenged these findings by exploring the role of practice within the direct and discovery frameworks. Their results suggested that direct instruction alone did not lead to acquisition of CVS for middle school students. Instead, varied practice over several weeks, irrespective of initially receiving instruction, produced deep, lasting learning. Strand-Cary and Klahr (2008) replicated Klahr and Nigam's findings and also found evidence of path independence of skill; those mastering CVS by the three month mark, irrespective of condition, significantly outperformed non-CVS masters on transfer tasks three years later.

Rather than viewing these results as evidence for or against direct instruction or discovery learning on the whole, they can be viewed in a different way. They can be viewed as evidence supporting that independent instructional tasks or components *in combination,* such as receiving an overview, practicing a skill repeatedly over time, or practicing in a concrete context, produces acquisition and retention of transferrable skill. Though not directly addressed in these studies, one such task present in all conditions yielding the highest learning gains was verbal or written self-explanation. In Klahr and Nigam's study, for example, students in their direct instruction condition critiqued experimental designs and provided justifications, a form of self-explanation (cf. Aleven & Koedinger, 2002), on why those designs were or were not confounded. We hypothesized that this self-explanation component played a role in long-term retention and transfer of CVS.

To test this hypothesis, we extended Klahr and Nigam's (2004) study by comparing the effectiveness of two types of direct instruction, with and without self-explanation (justification) tasks, versus discovery learning on the acquisition, transfer, and retention of CVS in the context of a virtual ramp environment. Our "direct" and "discovery" conditions represent slightly different approaches than these terms typically connote. Our direct instruction conditions were variants of guided inquiry in which students were taught CVS in a concrete context, the ramp environment. In contrast, our discovery learning condition was highly unguided; students attempted to construct unconfounded experiments receiving no feedback or instruction.

Previously, we showed that both direct conditions significantly outperformed the discovery condition at remediating multiply confounded experiments immediately after the intervention (Sao Pedro, Gobert, Heffernan & Beck, 2009). However, when retesting a subset of our sample six months later, direct-no explanation students' performance dropped sharply whereas the direct+explanation students maintained their skill (Sao Pedro, Gobert & Raziuddin, 2010). This provided evidence that prompting students to justify experimental designs in addition to our other direct instruction components supported deep learning (Chi & Bassok, 1989; Pirolli & Recker, 1994; Chi, 1996; Renkl, 1997; Hausmann & Chi 2002) and knowledge integration (Linn & Hsi, 2000) of CVS. Given these findings, we explored in this work if the self-explanation task, the key aspect of the direct+explain learning condition, directly and uniquely contributed to immediate and long-term skill at designing unconfounded experiments, describing the CVS procedure, and solving problems requiring CVS as compared to typical predictors such as pretest scores and time-on-task, and other learning condition-specific components.

Method

Participants

Participants were students from a public middle school in central Massachusetts whose students typically struggle with science. In 2008, 92% of eighth grade students at this school scored "below proficient" on the science MCAS standardized test (Massachusetts Department of Elementary and Secondary Education (ESE), 2008). In our original analyses, we omitted students on individual education programs (IEPs) and one student who used an incorrect web browser, leaving 133 participants, 74 seventh and 59 eighth grade students

Materials

We used the Science Assistments System (Gobert et al., 2007; Gobert et al., 2009), an extension of the ASSISTment System for math (Razzaq, et al., 2005) to host our materials, run randomized controlled experiments, and log all student interactions. Within Science Assistments, we developed a set of materials with a virtual ramp experimentation environment (Figure 1) in the spirit Klahr and Nigam's (2004) physical ramp apparatus to instruct and assess students on CVS. In this virtual environment, students constructed experiments to determine if any of four dichotomous variables affected how far a ball rolled down the ramps. The ramp had four changeable variables: *surface* (smooth or rough), *ball type* (golf or rubber), *steepness* (low or high), and *run length* (long or short).

Students interacted with our ramp environment differently than the physical materials in Klahr and Nigam (2004). In our environment, variable values were changed using combo boxes rather than adding and removing ramp pieces. More importantly, in their study students built ramp configurations starting from a blank slate. In our environment, a configuration was always prebuilt. In other words, there was always an initial condition students must change to create an unconfounded experiment. The experiment's initial state could be *unconfounded* (all variables are controlled), *singly confounded* (one variable is not controlled), *multiply confounded* (more than one variable is not controlled). It could also be *uncontrasted* (the target variable is unchanged). We chose to have students remediate experiments rather than construct them from a blank slate to dissuade students from constructing the same exact setups to demonstrate how they would design an unconfounded experiment for a given target variable.

The ramp environment was utilized in several ways. It was used to conduct authentic CVS performance assessments with the environment by asking students to remediate an initial experimental setup so that it would test if a given target variable affected the outcome, thereby applying CVS. Also, to emulate our learning conditions we designed a series of Science Assistments activities that utilized the ramp environment. In particular, each question required students to perform some combination of reading descriptions, designing and running experiments, answering multiple choice questions, and typing in answers to open response questions designed to be similar to thinking aloud. More details about this instruction are described in the Procedure.

Aside from the ramp environment, we also developed three transfer assessments within Science Assistments to tap into deeper understanding of CVS. The first transfer assessment was a set of multiple choice items that required CVS knowledge (referred to as CVS Inquiry items) to answer correctly. These asked students to identify an unconfounded experiment, identify a CVS procedure step, and determine an appropriate experimental setup with a valid control condition. The second transfer assessment asked two high-level open-response questions about the ramp regarding how to determine if any one or a set of variables affected how far the ball rolled (referred to as CVS Explanation items). Here, students needed to explain that one could systematically repeat CVS to find out how each variable affected the outcome. Specifically, they asked how (1) they "could determine if one particular variable affects how far the ball rolls in the ramp experiment" and (2) "when there are many variables that can be changes, explain how [they] could determine how each variable affects the distance the ball rolls." The final transfer assessment included a second CVS assessment battery composed of items developed by Strand-Cary and Klahr (2008). Both the CVS Inquiry and Explanation items were used in our immediate and delayed posttests while the Strand-Cary and Klahr items were administered only at the delayed posttest.

Procedure

In our procedure, outlined in Figure 2, students were first introduced to the ramp environment and pretested on their skill at constructing four unconfounded ramp experiments and answering our CVS Inquiry items. Science Assistments then randomly assigned each student to either the direct+explain, direct-no explain, or discovery condition. Within each condition, students practiced how to perform CVS in the context of the ramp environment. Immediately following the intervention, students were again tested on their skill at constructing unconfounded ramp setups, answering the same CVS Inquiry items as the pretest, and answering the CVS Explanation open-response items. Six months later, we administered a delayed posttest to the 7^{th} grade students. They again answered the same items as the immediate posttest in addition to the Strand-Cary and Klahr (2008) items.

Depending on the learning condition, students practiced CVS in different ways as highlighted in Figure 3. Here, we only describe the details of the direct+explain condition since our analyses focus specifically on this condition. In this condition, students were first asked to read an overview of CVS with examples of confounded and unconfounded ramp setups. After reading the overview, students addressed whether a series of six ramp setups had correctly controlled for variables. For each such practice task, students first responded to a yes/no probing question asking if a ramp configuration let them "tell for sure" if a variable affected how far a ball would roll. Next, they were given two opportunities to justify their choice by answering open response questions asking them to explain (type) their reasoning. The first open response question asked them simply to explain why they believed they could tell for sure if that variable affected the outcome. The second open response question presented the same ramp setup and allowed the students to run the experiment as many times as desired *without changing the setups.* After running the experiment, students were asked to again explain their reasoning. Finally, students received instruction by reading an explanation why the experiment was confounded or unconfounded for the target variable. For confounded setups, students were told which variables were confounded. For more information on this procedure or the other learning conditions, see (Sao Pedro et al., 2009).

Coding of Self-Explanations in the Direct+Explain Condition

As mentioned above, students had two opportunities to explain why an experimental setup was or was not confounded in each of six practice tasks in the direct+explain condition. Each explanation was scored out of 8 points according to the rubric described in Table 1. The rubric was designed to codify consistency with their initial evaluation of the setup and progressively deeper understanding of the procedural and conceptual components of CVS. Higher scoring responses reiterated their evaluation of the setup (probe question), identified the target variable, related the target to other variables, explained the CVS procedure, explained why the setup was unconfounded / confounded and identified which variables, if any, were confounded. Lower scoring responses either were minimal, irrelevant explanations or represented shallow understanding of CVS. For example, given an unconfounded experiment testing the effects of steepness, one student who wrote "the more steeper, the farther the ball will roll" received 4 points for their response (1 point each for rubric items A, B, C and E). This student provided a correct interpretation of the results of the unconfounded experiment and identified the target variable, but did not mention CVS. As another example, a student who answered "yes because everything is the same except the steepness" received a full score.

Results

We analyzed the degree to which the self-explanation tasks within the direct+explain condition uniquely predicted acquisition, transfer, and retention of CVS as measured by each type of CVS immediate and delayed posttest. These dependent measures included: (1) procedural CVS skill within our ramp environment, (2) CVS problem solving skill assessed with the CVS Inquiry items and Strand-Cary and Klahr (2008) items, and (3) skill at describing the CVS procedure, assessed by the CVS Explanation open response items[1]. The self-explanation task may not have been the only factor contributing to learning. For example, we include pretest variables, other intervention performance scores, and timing information since time-on-task has been found to reliably predict learning (e.g., Helmke & Renkl, 1992). We considered the following as potential predictors of performance:

- English Language Learner (ELL) status since the intervention required reading
- CVS Inquiry item pretest score
- Ramp pretest score
- Number of setups (out of 6) correctly identified as confounded/unconfounded
- Total time spent reading the initial instructional overview of CVS
- Total time spent considering if the setup was confounded or not during the probe question
- Total time spent reading explanations why each of the six setups was or was not confounded
- Total time spent writing self-explanations
- Quality of the self-explanations

We computed self-explanation quality by first tallying the scores for both self-explanation opportunities within each practice ramp setup (max 16 points per setup) and then taking the maximum score attained over all six setups. We used this metric as opposed to the cumulative or average score since students who understood CVS may not properly self-explain in later practice attempts due to their having felt they already demonstrated mastery of the skill. A cumulative or average score would have biased against these students. Additionally, three of the time-based variables--CVS overview reading time, practice opportunity explanation reading time, and self-explanation construction time--were transformed using the natural logarithm function to ensure normality.

Since we focus on the direct+explain learning condition, we considered only those 45 students randomly assigned to that condition in the analysis. We determined the impacts of our predictors through correlations and, where applicable, conducted forward multiple regressions on each immediate and delayed posttest measure. There were, however, imbalances in the number of students available out of those 45 to be used in predicting each posttest measure for two reasons. First, we only collected delayed performance metrics for our 7th grade participants meaning that only the 7th grade students could be used to analyze long-term retention. Second, some students did not complete the initial CVS inquiry pretest or ramp pretest since they were administered on separate days. For these reasons, we separately analyzed the effects of the direct+explain instructional components on each immediate and delayed posttest measure. We also used only those pretests we felt relevant for prediction of our dependent measures based on our prior work to preserve statistical power. Finally, like Klahr & Nigam (2004), students who scored perfectly on the ramp pretest were omitted to exclusively analyze the performance of students who have not mastered CVS.

Immediate Posttest Predictors

Correlations between each of the intervention components and immediate posttest measures, shown in Table 2, revealed several significant relationships. As expected, the correlation between the ramp pretest and immediate posttest, in which students remediated confounded experiments, was significant and moderate, $r(32)=.43$, $p=.013$. The correlation was weaker and marginal between the CVS inquiry pretest and posttest, $r(27)=.33$, $p=.085$. Time-on-task measures, though, were on the whole not correlated with posttest measures with the exception of the time spent reading the initial CVS tutorial. That predictor correlated with the open response posttest, $r(27)=.43$, $p=.022$. Additionally, the number of setups correctly identified as confounded or unconfounded in the initial probing question during the intervention was moderately correlated with open-response score, $r(27)=.45$, $p=.016$.

[1] When analyzing CVS Explanation performance, we used only students who explained procedures for constructing unconfounded experiments. Some students misunderstood the question and explained which variables affected how the ball rolled instead of explaining a procedure; they were treated as missing data. See Sao Pedro, Gobert & Raziuddin (2010) for more details on this and for information on how the open response item scores were computed.

Most importantly, self-explanation quality was the strongest predictor of each posttest among all the intervention components. It was strongly correlated with performance on the open response items asking students to explain the CVS procedure in their own words ($r(27)=.72$, $p<.001$), moderately correlated with the immediate ramp posttest ($r(32)=.46$, $p=.007$), and weakly but marginally correlated with performance on the CVS inquiry items ($r(27)=.36$, $p=.064$). These findings suggest that self-explanation quality was linked to the acquisition and deeper understanding of CVS.

Since two predictors significantly correlated with our immediate ramp posttest (ramp pretest and self-explanation quality) and three predictors significantly correlated with our open response posttest (number of setups correctly identified, CVS tutorial reading time, and self-explanation quality), we ran forward multiple regressions to determine the unique contributions of each instructional component, particularly self-explanation quality, on performance. As shown in Table 3, self-explanation quality and ramp pretest both provided unique contributions in predicting skill at remediating confounded experiments. Self-explanation score predicted R^2=21% ($F(1,32)=8.24$, $p=.007$) of the variance and ramp pretest score added an additional R^2=10% of explained variance, $F(2,30)=6.65$, $p=.004$. Though there were three significant correlates with our open response score in which students answered more conceptual CVS questions in their own words, only the self-explanation quality appeared as a unique predictor explaining a large proportion of variance for that dependent measure, R^2 =51%, $F(1,26)=17.94$, $p<.001$). Thus, the quality of self-explanations was the most uniquely predictive factor of all the intervention components, predicting above and beyond the pretests, the number correctly labeled setups, and the other timing variables considered.

Delayed Posttest Predictors

In our previous work, we found that students in both direct conditions, with and without the explanation component, were better at remediating multiply confounded experiments immediately after the intervention (Sao Pedro, et al., 2009). However, only the group who provided self-explanations maintained their skills six months later during another assessment (Sao Pedro, Gobert & Raziuddin, 2010). Therefore, we analyzed if self-explanation quality played a significant and unique role in predicting long-term retention of CVS.

Correlations, shown in Table 4, revealed a similar relationship pattern as the immediate posttest. Self-explanation quality played a pronounced role in predicting retention; it was the only significant predictor of performance at remediating confounded ramp experiments, $r(14)=.62$, R^2=38%, $p=.014$, and at answering the conceptual open-response questions, r(12)=.62, R^2= 38%, $p=.024$. It was also marginally correlated with the additional measure of CVS understanding at delayed posttest, the Strand-Cary and Klahr (2008) items, $r(15)=.45$, $p=.081$. In terms of pretest scores, only the CVS Inquiry pretest was predictive of its respective delayed posttest measure, $r(15)=.54$, $p=.032$. Surprisingly, several time-on-task variables were marginally and negatively correlated with the delayed CVS Inquiry posttest: time responding to the probing question ($r(15)=-.46$, $p=.072$), time reading descriptions why an experiment was or was not confounded ($r(15)=-.46$, $p=.076$), and time spent writing self-explanations ($r(15)=-.49$, $p=.057$). These time-on-task findings could indicate that students who were able to process the lesson more quickly were able to retain information about CVS in the long-term. Finally, none of the predictors significantly correlated with

In summary, the quality of self-explanations given by students was a strong and unique predictor of immediate acquisition and long-term retention of the procedural components of CVS relative to other intervention-specific components, time-on-task variables, and pretest scores. This was evidenced by strong correlations and larger percentages of unique variance explained in predicting immediate and delayed skill at remediating confounded experiments. Transfer results were slightly mixed, however. Self-explanations quality was highly predictive of immediate and delayed skill at answering high-level conceptual questions requiring articulation of CVS, thus indicating a degree of transfer. But, it only marginally correlated with immediate and delayed CVS problem solving skill.

Discussion and Conclusions

Although there is currently a large body of research on inquiry learning, there are debates on which pedagogical strategies produce deep understanding, particularly with respect to discovery learning and direct instruction (c.f. Kirschner, Sweller, & Clark, 2006; Hmelo-Silver, Duncan, & Chinn, 2007). Several researchers have argued in favor of using direct instruction over discovery learning when teaching the control of variables strategy (CVS) (Klahr & Nigam, 2004; Strand-Cary & Klahr, 2008) while others have argued the opposite (e.g. Dean Jr. & Kuhn, 2007). Rather than vying for an all-in-one approach, others have studied which instructional components led to differences in learning. For example, Triona and Klahr (2003) found no differences between students learning CVS using physical or virtual materials with a human tutor. Zohar and David (2008) found that adding meta-strategic knowledge instruction brought lower prior knowledge students up to the levels of higher achieving students when teaching CVS in a classroom setting. Lorch et al. (2010) found that instruction of CVS taught students how to handle confounds whereas instruction that let students manipulate experimental setups led to better detection of identifying valid and non-contrastive experimental designs. Finally, Siler, Klahr, Magaro, Willows & Mowrey (2010) found no performance differences between students who were prompted to think about causality inferences that could be made from experimental setups, the "logic" of CVS, and those who were not prompted in a computerized CVS lesson, similar to our learning environment.

In a similar vein, we focused on one instructional component, self-explanation in the form of justifications (cf. Aleven & Koedinger, 2002), and analyzed if it was a necessary component for yielding deep, long-term understanding of CVS. In previous work, we found that students in two direct instruction conditions, with and without explanation, were better at correcting multiply confounded setups than a discovery condition at immediate posttest (Sao Pedro, et al., 2009). When retesting a subset of our sample six months later, we found that the direct+explanation condition was the only condition showing overall immediate and lasting retention of skill (Sao Pedro, Gobert & Raziuddin, 2010). Given these overall effects, we explored which factors within the direct+explain condition would most strongly predict deep, lasting learning. We found that the quality of self-explanations in the form of justifications why an experimental setup was or was not confounded during practice was highly predictive of immediate and long-term skill at constructing unconfounded experiments (near-transfer) and explaining the CVS procedure (far-transfer). The strength of its prediction was beyond other learning condition-specific components and typical posttest predictors, namely time-on-task variables, and pretest scores. Our transfer results were slightly mixed, however, since quality of self-explanation only marginally correlated with immediate and delayed CVS problem solving skill.

Amalgamating our previous and current findings suggests that it is important to engage and support students in constructing quality self-explanations not only to maximize learning, as has been found in previous studies (e.g. Chi, DeLeeuw, Chiu, & LaVancher, 1994; Renkl, 1997), but also to foster retention of knowledge. Though our results suggest that self-explanation alone is necessary to foster acquisition, long-term retention and a degree of transfer of CVS, we acknowledge that the other direct instruction components may still be required as well since we did not test the efficacy of a self-explanation learning condition without direct instruction (i.e. discovery+explain). Dean Jr. and Kuhn's (2006) results, however, suggest that such direct instruction may not be required. They found that practicing CVS regularly and working in dyads, in addition to providing explanations, led to learning. Dean Jr. and Kuhn's findings do add evidence to our argument, but a randomized controlled study is still needed to test this idea more conclusively.

Given the importance of quality self-explanations, we should scaffold students in constructing these explanations. The manner in which we scaffold is still an open question. We could provide "good" and "bad" CVS self-explanation examples to model reasoning (McNeill & Krajcik, 2008). We could also initially tightly scaffold explanations using a fill-in-the-blanks approach (McNeill & Krajcik, 2008) or template structures (Aleven & Koedinger, 2002) and then fade gradually to open-responses (McNeill, Lizotte, Krajcik, & Marx, 2004) as students' explanation skills progress. Such additions may help students acquire and maintain understanding of CVS.

As previously stated, there is a debate in the science education community which has juxtaposed direct instruction with open-ended discovery. We recognize that inquiry approaches range from more to less open-ended in terms of structure and believe that different combinations of tasks from each paradigm, such as initial instruction or repeated, varied practice can play different roles in inquiry skill acquisition. Thus, we argue that it is not a question about whether direct instruction is better than discovery learning, but rather a question about what tasks or aspects from each paradigm consistently yield robust learning. In our study, we identified one such task for acquiring and retaining robust understanding of CVS, having students provide quality self-explanations. Thus, scaffolding should be available to support students who do not provide good explanations. Explanation tasks may also be beneficial for other inquiry skills, such as hypothesizing or analyzing, and should also be scaffolded. The degree of scaffolding needed and the robustness of learning yielded are all empirical issues addressed in our work (Gobert et al., 2007; Gobert et al., 2009).

Acknowledgments

This research was funded by the National Science Foundation (NSF-DRL#0733286; NSF-DGE# 0742503) and the U.S. Department of Education (R305A090170). Any opinions expressed are those of the authors and do not necessarily reflect those of the funding agencies.

References

Aleven, V., & Koedinger, K. (2002). An Effective Metacognitive Strategy: Learning by Doing and Explaining with a Computer-based Cognitive Tutor. *Cognitive Science , 26*, 147-179.

Aleven, V., & Koedinger, K. (2000). The Need for Tutorial Dialog to Support Self-Explanation. In C. Rose, & R. Freedman, *Building Dialogue Systems for Tutorial Applications. Papers from the 2000 AAAI Fall Symposium* (pp. 65-73). Menlo Park, CA: AAAI Press.

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L.B. Resnick (Ed.). In L. R. (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361-392). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chen, Z., & Klahr, D. (1999). All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy. *Child Development , 70* (5), 1098-1120.

Chi, M. (2000). Self-explaining Expository Texts: The Dual Process of Generating Inferences and Repairing Mental Models. In R. Glaser, *Advances in Instructional Psychology* (pp. 161-238). Mahweh, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. (1996). Constructing Self-Explanations and Scaffolded Explanations in Tutoring. *Applied Cognitive Psychology, Vol. 10, S33-49* .

Chi, M., & Bassok, M. (1989). Learning from Examples Via Self-Explanations. In L. Resnik, *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser* (pp. 251-282). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M., DeLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science , 18*, 439-477.

Dean Jr., D., & Kuhn, D. (2006). Direct Instruction vs. Discovery: The Long View. *Science Education* , 384-397.

Gobert, J., Heffernan, N., Ruiz, C., & Kim, R. (2008). *Assistments Meets Inquiry. Annual Report submitted to the National Science Foundation for NSF-DRL# 0733286.*

Gobert, Janice (Principal Investigator); Heffernan, Neil; Koedinger, Ken; Beck, Joseph (Co-Principal Investigators). (2009). ASSISTments Meets Science Learning (AMSL; R305A090170). Awarded February 1, 2009 from the U.S. Dept. of Education.

Gobert, Janice (Principal Investigator); Heffernan, Neil; Ruiz, Carolina; Kim, Ryung (Co-Principal Investigators). (2007). AMI: ASSISTments Meets Inquiry (NSF-DRL# 0733286). Awarded September 2007 from the National Science Foundation.

Hausmann, R. G., & Chi, M. T. (2002). Can a Computer Interface Support Self-Explaining? *Cognitive Technology , 7* (1), 4-14.

Helmke, A., & Renkl, A. (1992). The Munich Attention Inventory (MAI): An Instrument for the Systematic Observation of Students' On-Task Behavior during Instruction. *Diagnostica* , 130-141.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Krischner, Sweller, and Clark (2006). *Educational Psychologist, 42(2)* , 99-107.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41(2)* , 75-86.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning. *Psychological Science, 15(10)* , 661-667.

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

Kuhn, D. (2005). What needs to be mastered in mastery of scientific method? *Psychological Science, 16(11)* , 873-874.

Linn, M. C., & Hsi, S. (2000). *Computers, Teachers, Peers: Science Learning Partners.* Mahwah, NJ: Erlbaum Associates.

Lorch Jr., R. F., Lorch, E. P., Dunlap, E. E., Hodell, E. C., & Dunham, F. B. (2010). Learning the Controlf of Variables Strategy in Higher and Lower Achieving Classrooms: Contributions of Explicit Instruction and Experimentation. *Journal of Educational Psychology , 102* (1), 90-101.

Massachusetts Department of Elementary and Secondary Education (ESE). (2008). Retrieved 10 2008, from http://profiles.doe.mass.edu

McNeill, K., & Krajcik, J. (2008). Inquiry and Scientific Explanations: Helping Students Use Evidence and Reasoning. In J. Luft, R. Bell, & J. Gess-Newsome, *Science as Inquiry in the Secondary Setting* (pp. 121-134). Arlington, VA: National Science Teachers Association Press.

McNeill, K., Lizotte, D., Krajcik, J., & Marx, R. (2004, April). Supporting Students' Construction of Scientific Explanations Using Scaffolded Curriculum Materials and Assessments. *Paper presented at the Annual Conference of the American Education Research Association, San Diego.*

National Committee on Science Education Standards and Assessment, National Research Council . (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

Pirolli, P., & Recker, M. (1994). Learning Strategies and Transfer in the Domain of Programming. *Cognition and Instruction , 12*, 235-275.

Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N., Koedinger, K. R., Junker, B., et al. (2005). The Assistment Project: Blending Assessment and Assisting. In G. M. C.K. Looi, *Proceedings of the 12th International Conference on Artificial Intelligence In Education* (pp. 555-562). Amsterdam: ISO Press.

Renkl. (1997). Learning from Worked-out Examples: A Study on Individual Differences. *Cognitive Science , 21* (1), 1-29.

Sao Pedro, M. A., Gobert, J. D., & Raziuddin, J. (2010). Comparing Pedagogical Approaches for the Acquisition and Long-Term Robustness of the Control of Variables Strategy. In K. Gomez, L. Lyons, & J. Radinsky (Ed.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers* (pp. 1024-1031). Chicago, IL: International Society of the Learning Sciences.

Sao Pedro, M., Gobert, J., Heffernan, N., & Beck, J. (2009). Comparing Pedagogical Approaches for Teaching the Control of Variables Strategy. *N.A. Taatgen & H. vanRijn (Eds.), Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (pp. 1294-1299). Austin, TX: Cognitive Science Society.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49* , 31-57.

Siler, S., Klahr, D., Magaro, C., Willows, K., & Mowery, D. (2010). Predictors of Transfer of Experimental Design Skills in Elementary and Middle School Children. *Intelligent Tutoring Systems, Part II* (pp. 198-208). Berlin: Springer-Verlag.

Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills; Instructional effectiveness and path independence. *Cognitive Development 23* , 488-511.

Triona, L. M., & Klahr, D. (2003). Point and Click or Grab and Heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition & Instruction, 21* , 149-173.

Zohar, A., & David, A. B. (2008). Explicit Teaching of Meta-Strategic Knowledge in Authentic Classroom Situations. *Metacognition Learning , 3*, 59-82.

Appendix: Figures and Tables

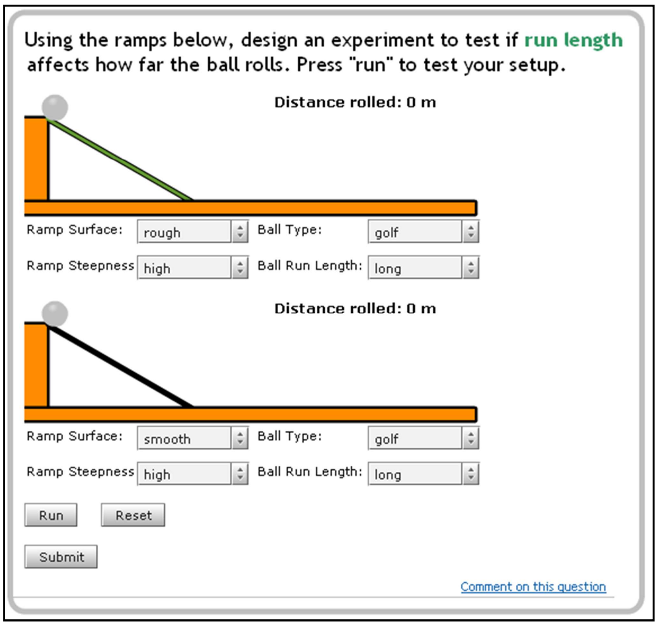

Figure 1. Science Assistments question using the ramp environment. Students constructed experiments to determine if any of four dichotomous independent variables: *surface*, *ball type*, *steepness*, and *run length* affected how far a ball rolls down the ramps. Initially, ramp setups could be *unconfounded* (all variables are controlled), *singly confounded* (one variable is not controlled), *multiply confounded* (more than one variable is not controlled), and/or *uncontrasted* (the target variable is unchanged). The setup shown above is uncontrasted and singly confounded because the target variable, run length, is the same for each setup and one extraneous variable, surface, is not controlled.

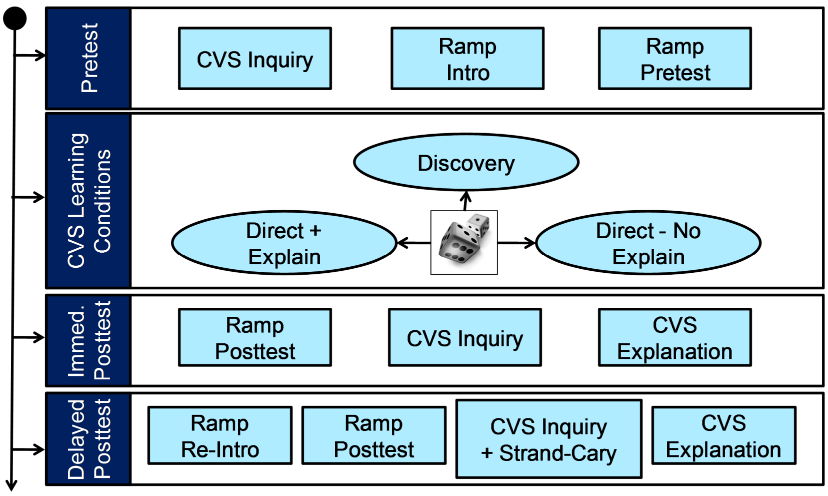

Figure 2. Experimental Design for this study. Students were randomly assigned to one of three learning conditions, Direct+Explain, Direct-No Explain or Discovery.

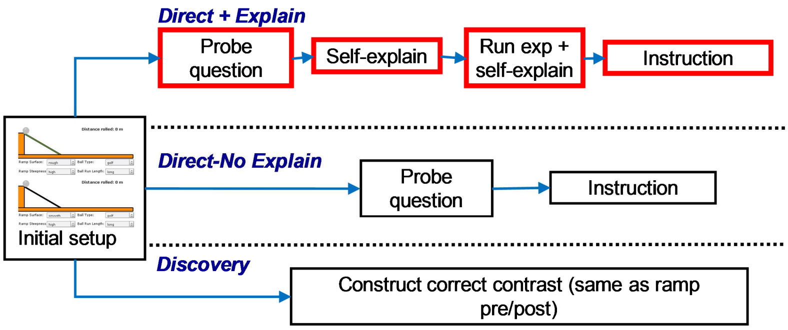

Figure 3. Activity sequence for each learning condition: students in all conditions practiced CVS on six different initial setups, some confounded, and some unconfounded. Students in both direct conditions first read an introduction about CVS before practicing. The analyses in this work focus on the behaviors exhibited and self-explanations written by students in the Direct+Explain condition (highlighted in red).

Table 1: Direct instruction + explanation self-explanation scoring rubric. This rubric was used to score both self-explanations (self-explanation and run experiment + self-explanation in Figure 3) provided by students for each practice loop.

	Rubric Element	Max Score
A	Is this student's response meaningful or did they simply write jibberish or leave the answer blank? Did the student respond "I don't know" or something else indicated they did not understand the question or how to answer it?	1
B	Is their explanation in agreement with their initial "yes/no" prediction? For example, if the student answered in the probe question that the ramps were not confounded, did they also state in their explanation that they were not confounded?	1
C	Is their explanation of "telling for sure" correct given the ramp setup? Did they either correctly identify if the setup was confounded or not (irrespective of understanding CVS) or provide an interpretation about the outcome consistent with the setup?	1
D	Do they state anything about the CVS procedure in general, even if their explanation is incorrect given the ramp setup?	1
E	Did they correctly identify the target variable?	1
F	Did they correctly relate the target to the other variables? Do they say that the target is what needs to be changed to make a comparison and other variables need to be held constant?	1
G	If variables were confounded, did they correctly identify which variables were confounded? If the variables were not confounded, did they correctly identify that 'all other variables except the target' were controlled? If the setup is multiply confounded, give full score only if all confounded variables are mentioned.	2

Table 2: Correlations between direct+explain instructional components and immediate posttest measures.

	Immediate Post Tests		
	Ramp	*CVS Inquiry*	*CVS Open Response*
	N =33	*N* =28	*N* =28
Predictors			
1. ELL Status[a]	-.30#	-.10	-.09
2. CVS Inquiry pretest	----	.33#	----
3. Ramp pretest	.43*	----	.27
4. Intervention: number of correct setups identified	.32	.22	.45*
5. Intervention: Ln of CVS overview reading time	.29	.04	.43*
6. Intervention: time responding to probe question	-.12	-.13	.01
7. Intervention: Ln of practice opportunity explanation reading time	.00	-.13	.21
8. Intervention: Ln of self-explanation construction time	-.08	-.18	.02
9. Intervention: Self-explanation quality	.46**	.36#	.72***

#$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$

[a]Point-biserial correlation coefficient reported

Table 3: Predictors entered at each step in the forward linear regression when predicting each immediate posttest measure.

Immediate Posttest Measure	*N*	Step	Predictor Added	R^2	ΔR^2	*F*
Ramp	33	1	Self-Explanation Quality	.21	.21	8.24**
		2	Ramp Pretest	.31	.10	6.65**
CVS Explanation	28	1	Self-Explanation Quality	.51	.51	27.35***

**p < .01; ***p < .001

Table 4: Correlations between direct+explain instructional components and delayed posttest measures.

	Delayed Post Tests			
	Ramp	*CVS Inquiry*	*CVS Open Response*	*Strand-Cary & Klahr*
	N =15	*N* =16	*N* =13	*N* =16
Predictors				
1. ELL Status[a]	-.51#	-.30	-.06	-.33
2. CVS Inquiry pretest	----	.54*	----	----
3. Ramp pretest	-.01	----	-.29	----
4. Intervention: number of correct setups identified	.07	.05	.30	.08
5. Intervention: Ln of CVS overview reading time	-.01	-.08	-.01	.21
6. Intervention: time responding to probe question	.23	-.46#	.10	.12
7. Intervention: Ln of practice opportunity explanation reading time	-.06	-.46#	.10	.07
8. Intervention: Ln of self-explanation construction time	-.11	-.49#	-.10	-.02
9. Intervention: Self-explanation quality	.62*	.33	.62*	.45#

#p < .10; *p < .05

[a]Point-biserial correlation coefficient reported