Comparing Pedagogical Approaches for the Acquisition and Long-Term Robustness of the Control of Variables Strategy

Michael A. Sao Pedro, Janice D. Gobert, Juelaila J. Raziuddin
Worcester Polytechnic Institute
100 Institute Rd. Worcester, MA 01609
Email: mikesp@wpi.edu, jgobert@wpi.edu, juelaila@wpi.edu

Abstract: This study compared three pedagogical approaches on the acquisition and robust understanding of the control of variables strategy (CVS). In Sao Pedro et al. (2009), we showed that two direct learning conditions (with and without reification) significantly outperformed the discovery learning condition for constructing unconfounded experiments starting from an initially multiply confounded experimental setup. In the study described here, we retested 57 students six months later on constructing unconfounded experiments in a virtual ramp environment, solving problems requiring CVS, and explaining CVS. Collapsing over time, we found that the direct+reify condition had more robust learning than either the direct-no reify or discovery learning conditions on constructing unconfounded experiments. At the delayed posttest, we found a strong trend favoring the direct+reify condition over the other conditions as measured by tasks designing unconfounded experiments starting from a multiply confounded state.

Introduction

Currently, there is a debate in the science education community regarding the effectiveness of discovery vs. direct instruction. Critics of discovery learning claim that it may be less effective when compared to instructional approaches with emphasis on guidance (Kirschner, Sweller, & Clark, 2006). In particular, it has been found that during open-ended inquiry, students can have many false starts (Schauble, 1990), and have difficulty designing effective experiments, forming testable hypotheses, adequately monitoring what they do (de Jong, 2006), linking hypotheses and data, and drawing correct conclusions (Klahr & Dunbar, 1988; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995). On the other hand, direct instruction, in its traditional form, has also received criticism, namely, that it can turn into rote instruction leading students to lose engagement and develop inert knowledge that cannot be flexibly applied or transferred (Bereiter & Scardamalia, 1989).

Within these approaches, researchers have begun to determine if these instructional paradigms yield successful transfer of knowledge and skills. Some have shown that explicitly teaching *strategies and skills* in one context can successfully transfer to another context (Klahr & Nigam 2004; VanLehn et al., 2005) or domain (Chi & VanLehn, 2007). More specifically and relevant to the present study, Klahr and Nigam (2004) found that students who were taught the control of variables strategy (CVS) in a direct learning condition significantly outperformed those in a discovery learning condition on a near-transfer test of CVS. Furthermore, those who mastered CVS, irrespective of learning condition, outperformed non-masters on a near-transfer test of this skill. These results suggest that the purported benefits of discovery learning, particularly the deeper learning, may not always occur. These Klahr and Nigam (2004) findings, though, are not without critique (cf., Hmelo-Silver, Duncan, & Chinn, 2007). Others such as Kuhn (2005) have criticized Klahr and Nigam because they did not test students' knowledge about when and why to use CVS.

In a similar vein, others have explored the degree to which these instructional paradigms lead to long-term robustness of skills. For example, Strand-Cary and Klahr (2008) compared the long-term effectiveness of guided and unguided approaches to teaching CVS by repeatedly testing students over a much longer period of time, up to 3 years after their initial intervention. They found significant differences in skill at constructing unconfounded experiments favoring their explicit instruction condition immediately after their intervention, but that 3 months later, those in their exploration condition caught up to the explicit instruction group on average. Furthermore, those who mastered CVS by the 3 month mark, irrespective of condition, significantly outperformed non-CVS masters on transfer tasks 3 years later. Dean and Kuhn (2006) explored the role of practice and engagement within the direct and discovery frameworks on robust acquisition of CVS. They found that direct instruction alone did not lead to robust acquisition; practice and engagement, irrespective of initially receiving instruction, produced deep, lasting learning. In our work, we explored the role that self-explanation played in long-term retention and transfer of CVS skill.

Our first study, Sao Pedro, Gobert, Heffernan, & Beck (2009), an extension to Klahr and Nigam's (2004) study, compared the effectiveness of two types of direct instruction versus discovery with regard to the acquisition and transfer of CVS. We hypothesized that the self-explanation component in Klahr and Nigam's direct instruction condition could have played a role in students' acquisition of CVS. By adding our third condition, direct no-reify, which removed prompting for student explanations, we empirically tested if the explanations in the direct+reify condition affected the acquisition of CVS. The terms "direct" and "discovery" have slightly different meanings than is reflected in our learning conditions; that is, these terms typically represent polar opposites in terms of level of *directedness* given to students. Our direct instruction conditions portray variants of guided inquiry in which students are taught the procedure of CVS in a concrete context, a virtual ramp environment. Our results showed that in an immediate posttest following the interventions, students in both direct conditions significantly outperformed students in the discovery condition on tasks involving creating unconfounded experiments starting from an initally multiply confounded setup. The direct conditions, however, did not significantly differ from each other.

In the present study, we tested the efficacy of the reification task (self-explanation) on the acquisition and robust understanding of CVS over time, 6 months after the original Sao Pedro et al. (2009) study by administering a delayed posttest on participants from our original study. We compared the conditions on the understanding of CVS as evidenced by their skills at constructing unconfounded ramp setups, explaining the CVS procedure, and solving multiple choice and open response problems requiring CVS. Self-explanation has been found to support deep learning (Chi M. T., 1996) and knowledge intgegration (Linn & Hsi, 2000). Thus, in this study, we hypothesized participants in the direct+reify group would outperform the other conditions even though the direct-no reify group showed the same performance as direct+reify for certain items at the immediate posttest (Sao Pedro et al, 2009).

Method

Participants

Participants were seventh-grade students from a public middle school in central Massachusetts. We chose this group because middle school may be the time to optimally learn model-based inquiry skills (Schunn, Raghavan, & Cho, 2007). Participants all had the same science teacher through the school year who taught five heterogeneously grouped sections. Students at this school typically struggle with science. For example, in 2008, 92% of eighth grade students at this school scored below proficient on the science MCAS exam, a standardized test administered by the state (Massachusetts Department of Elementary and Secondary Education (ESE), 2008).

All class sections participated in the experiment. For this paper, we concerned ourselves only with the performance of those students who completed our initial intervention and posttest (Sao Pedro et al, 2009) as well as the delayed posttest. Further, we did not include data from students on individual education programs (IEPs) and one additional student who used an incorrect web browser during the initial intervention, leaving a total of 57 students in the analyses for this study.

Materials

We used The Science ASSISTment System (Gobert et al., 2007; Gobert et al., 2009), a web-based intelligent tutoring system, to host our materials, run randomized controlled experiments and log time-stamped student interactions with the system. We used several types of activities to assess students' acquisition of CVS. Students practiced authentic inquiry by constructing unconfounded experiments using our virtual ramp environment. The same ramp activity acted as a near-transfer performance assessment. We also developed a far transfer test involving several multiple choice and open response questions. Some were designed by us and others were obtained from a study conducted by Strand-Cary and Klahr (2008).

Ramp Environment

The ramp environment (Figure 1) was developed using the OpenLaszlo framework (www.openlaszlo.org). We created different kinds of questions by embedding the ramp environment within the ASSISTment System. The ramp apparatus has four variables that can be manipulated: *surface* (smooth or rough), *ball type* (golf or rubber), *steepness* (low or high), and *run length* (long or short). The objective is to set up the ramps so that the target variable is contrasted and all other variables are held constant. Pressing the "run" button causes each ball to roll down its respective ramp. Depending on each ramp's settings, the ball will roll different distances. Participants submit their final setup using the "submit" button. Pressing "reset" causes the balls to be placed back on the ramp and clears the distances rolled. On each press of the "run" or "submit" button, student information is logged. This includes a timestamp of the run or submission, the correctness of their setup in terms of CVS, and the current and previous ramp value settings.

Unlike the physical and virtual environments used in Klahr and Nigam (2004) and Strand-Cary and Klahr (2008) in which participants set up ramp conditions starting from a blank slate, our ramp setup was always initially set. There was always an initial condition students must change in order to create an unconfounded experiment. In a sense, this activity can be viewed as an experimental setup repair task since students must morph a given setup into one that follows CVS. The experiment's starting state could initially 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). Our instruction and feedback format was also different from previous studies; the CVS explanations given to the direct conditions were entirely computer-based whereas the other studies both used human tutoring.

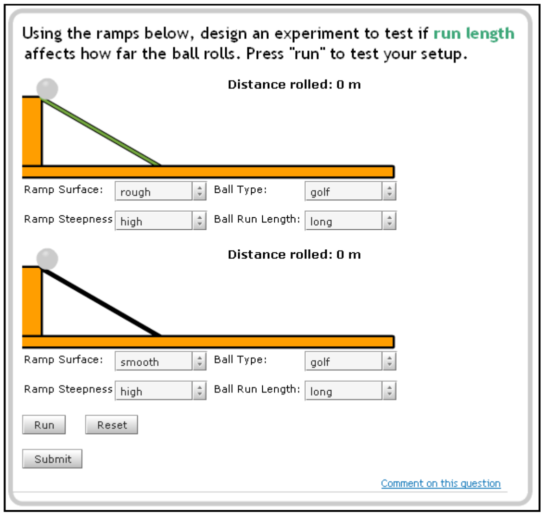

Figure 1. ASSISTment question using the ramp environment. This setup is uncontrasted and singly confounded because the target variable, run length, is the same for each setup and one extraneous variable, surface, is not controlled.

Multiple Choice and Open Response Transfer Items

In Sao Pedro et al (2009), we developed three multiple choice and two open response items to assess students' skills in using CVS to solve problems and in communicating the procedure. We reused these exact items again for this study. The multiple choice items, referred to as the WPI CVS Inquiry items, asked students to identify an unconfounded experiment, identify a CVS procedure step, and determine an appropriate experimental setup with a valid control condition. The two open response items asked students to describe in their own words the CVS procedure. 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." Students needed to clearly state that CVS was a procedure one could systematically repeat to find out how each variable affected the outcome.

For this study, we also added two multiple choice items and one open response item from a transfer test developed by Strand-Cary and Klahr (2008). Their multiple choice items on plant and seed growth required students to choose the correct experimental setup that contrasted a variable of interest and kept all others constant. Their open response item asked, without specific prompting for CVS usage, for students to critique a confounded experimental design that tested if beetles preferred to live in sun or shade.

Procedure

This study determined the long-term effects of the differing interventions on CVS acquisition using both the original posttest materials and the new Strand-Cary and Klahr (2008) assessments. In our original October 2008 experiment, students were first pretested on their skillfulness at creating unconfounded experiments in the ramp environment without receiving feedback. We also tested them on the WPI CVS Inquiry items, multiple choice questions involving CVS. The ASSISTment System then randomly assigned each student to either the direct+reify, direct-no reify, or the discovery condition. In each condition, students practiced how to perform CVS in the context of the ramp environment. Following the intervention, students were tested on their skill at constructing unconfounded ramp setups, answering the same WPI CVS multiple choice questions as the pretest, and explaining in their own words the CVS procedure within the context of the ramp environment.

In both direct conditions, students were first asked to read an overview of CVS with examples of confounded and unconfounded ramp setups. In this overview, we did not explain that CVS could be used to solve problems in other domains. After reading the overview, students addressed whether a series of ramp setups had correctly controlled for variables. In the *direct+reify* condition, participants first responded if a given ramp configuration enabled them "tell for sure" if a variable affected how far a ball would roll by responding "yes" or "no". Next, they answered an open response question asking them to explain their reasoning. For the same ramp setup, students then were allowed to run the experiment as many times as desired *without changing* *the setups* and explained again if they could tell for sure that target variable affected how far the ball would roll. Finally, students read an explanation why the experiment was confounded or unconfounded for the target variable. If the setup was confounded, students were told which variables were confounded. Students in the *direct-no reify* condition followed the exact same procedure, except they were not asked the two open response questions. Students in the *discovery* condition were instructed to create experiments that tested if a particular variable affected how far the ball rolled. The discovery condition students were neither given the initial CVS explanation nor any feedback about the correctness of their experimental setups. All conditions practiced on six identical initial ramp configurations. For more details on the original experimental procedure, see Sao Pedro, Gobert, Heffernan, & Beck, (2009).

The present study took place six months later in June 2009 and acted as a delayed posttest of CVS skill. Throughout the school year, students used the ASSISTment system extensively for math class but not for science-related activities. All students in each class section took the delayed posttest irrespective of having participated in the original experiment, though we only analyze those students for whom we have posttest data from our October 2008 experiment. Students first answered the WPI CVS Inquiry multiple-choice questions developed in the previous study. Next, students answered the subset of Strand-Cary and Klahr (2008) questions, including one open response question. Students were then presented a reintroduction to the ramp environment that described the ramp variables and how to interact with the simulation. Students then constructed four unconfounded ramp setups with different target variables. The initial setups were identical to those in the previous study's posttest. Finally, students again answered the open response items in which they explained CVS in the context of the ramp experiment.

Scoring

Multiple-choice questions were automatically scored by the ASSISTment system with a 1 if correct or 0 if incorrect. Correct ramp setups demonstrating CVS for the given target variable were awarded 1 point, 0 otherwise. All open response items were hand scored by two different graders according to a rubric. The Strand-Cary and Klahr (2008) open response question was graded out of 3 points, one point for identifying that the experimental setup was incorrect, and up to 2 points for correctly explaining the experiment's design flaws. Maximum points were only awarded if the student addressed the lack of control in the question's experimental design.

The two ramp open response items were combined into one measure to capture skill in correctly describing the CVS procedure, marked out of 3 points. Higher scores indicated deeper understanding of CVS. One point was given for identifying independent and dependent variables and the presence of some relationship between them. Two points were awarded for stating that the target variable needed to be contrasted and all other variables should be the same. Three points were given for understanding that CVS should be repeated for each variable to determine how each individually affected the outcome. For example, a student who received full credit answered as follows: "For each variable keep the others the same and change one variable at a time." When reviewing student responses, we realized a large number of students described how each specific variable affected the distance rolled instead of describing a procedure. If students answered with such evidential knowledge, they were scored according to a different rubric and were treated as missing data for this metric. Students falling into neither category, such as those responding "I don't know" received 0 points.

Results

We analyzed 57 students' immediate and delayed posttest scores to determine which condition(s) yielded better performance on each CVS measurement: problem solving using CVS via the multiple choice items, articulation and explanation of CVS via the ramp open response items, and authentic inquiry performance with the ramp environment. In particular, we determine if self-explanation in the direct+reify condition led to better performance over time for the various near-transfer and far-transfer CVS tasks. We also analyzed if the findings favoring the direct+reify and direct-no reify groups over the discovery group on multiply confounded ramp items were robust for the delayed posttest. For each analysis of variance below, we report partial η^2 values as measurements for effect size.

Differences between Groups on Transfer Items

In these analyses we looked for trend differences between groups at immediate and delayed posttest for the WPI CVS inquiry test and the ramp open response items. We also compared groups' performance on the Strand-Cary and Klahr (2008) items. Means and standard deviations for all transfer items are presented in Table 1. Only students who completed all the questions for a measurement were included in that measurement's analysis. We conducted a repeated measures ANCOVA to compare performance on the WPI CVS Inquiry multiple choice items over time. The WPI CVS Inquiry pretest was used as a covariate. Though time was not a significant within-subjects factor (Wilks λ=1.00, $F(1,48)=0.19$, $p=.668$, partial $\eta^2=.004$), the interaction effect between time and condition was marginally significant, Wilks λ=.907, $F(2,48)=2.47$, $p=.095$, partial $\eta^2=.093$, suggesting

Table 1: Means and standard deviations for the transfer items at pretest (t0), immediate posttest (t1), and delayed posttest (t2).

		Direct+Reify			*Direct-No Reify*			*Discovery*		
	Max	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
WPI CVS MultCh t0	3	1.38	1.07	21	1.00	0.85	15	1.06	1.00	16
WPI CVS MultCh t1	3	1.81	0.91	22	1.25	1.00	16	1.52	1.07	19
WPI CVS MultCh t2	3	1.36	1.00	22	1.25	0.93	16	1.74	0.93	19
Ramp CVS OpenR t1*	3	1.06	1.11	18	1.57	0.79	7	1.08	1.12	13
Ramp CVS OpenR t2*	3	1.11	1.08	18	0.80	0.92	10	0.80	0.79	10
Strand-Cary & Klahr t2	3	1.38	1.00	21	1.38	0.74	16	1.61	0.96	18

* Open response items include only those who responded with procedures, not evidential statements.

there may have been performance differences between the conditions as time progressed. As Table 1 indicates, performance in the direct-no reify condition was the same for the each posttest, M=1.25. Students in the direct+reify condition showed better performance on average at the immediate posttest, but did not maintain those scores. The discovery condition's performance increased on average as time progressed. Condition was not a significant between-subjects factor, $F(2,48)=1.78$, $p=.179$, partial $\eta^2=.069$. This indicated no performance differences between the groups when collapsing over time.

Another repeated measures ANCOVA determined if any intervention affected students' skill in explaining CVS. In this analysis, we compared only students who explicated procedures for constructing unconfounded experiments in response to the two ramp open response questions since some students explained which variables affected how the ball rolled instead of explaining a procedure. There was neither a significant effect for time (Wilks λ=.98, $F(1,21)=0.49$, $p=.492$, partial $\eta^2=.023$) nor for the interaction between time and condition, Wilks λ=.93, $F(2,21)=0.81$, $p=.460$, partial $\eta^2=.071$. Condition was also not a significant between-subjects factor, $F(2,21)=0.23$, $p=.800$, partial $\eta^2=.021$.

Finally, we analyzed if there were differences between the groups on the Strand-Cary and Klahr (2008) items using an ANOVA. We normalized the beetles open response item in order to equally weight all items for the dependent measure. No significant difference was found between the groups on this measure, $F(2,52) = 0.39$, $p=.681$, partial $\eta^2=.015$.

Taking these transfer results as a whole, we saw no significant differences in trends and no significant differences between conditions for any of the transfer item subsets. This suggests that our different learning conditions did not enable transfer from constructing unconfounded ramp experiments to solving more complex reasoning problems requiring CVS. As for explaining CVS, we believe we would expect to see differences between learning conditions if the questions were worded such that students answered with procedures rather than evidential statements. We believe this because there were significant differences between groups' overall skills in constructing unconfounded experiments in the ramp environment as discussed next.

Comparison of Ramp Items

We compared each condition's skill in constructing unconfounded ramp experiments over time using a repeated measures ANCOVA. Total ramp scores at each time were dependent measures and total ramp pretest score was used as a covariate. Means and standard errors for ramp performance over time are shown in Figure 2. Four students did not take the original ramp pretest and were not included in this analysis, leaving 53 students. Within-subjects tests revealed no significant main effect for time, Wilks λ=.97, $F(1,49)=1.36$, $p=.249$, partial $\eta^2=.027$, and no significant interaction between time and condition, Wilks λ=.92, $F(2,49)=2.03$, $p=.143$, partial $\eta^2=.076$. However, condition was a significant between-subjects factor, $F(2,49)=3.26$, $p=.047$, partial $\eta^2=.117$, controlling for ramp pretest score. Post hoc comparisons revealed that the direct+reify condition constructed significantly more experiments adhering to CVS than the discovery condition (M=0.92, SE=0.43, $p=.037$, 95% CI=[0.06, 1.79]) and the direct-no reify condition (M=0.98, SE=0.45, $p=.033$, 95% CI=[0.84, 1.88]). No significant difference was found between the direct-no reify and discovery conditions, $p=.903$.

Since there was a significant between-subjects effect on condition, we also analyzed group differences solely at the time-2 delayed posttest using an ANCOVA with ramp pretest score as a covariate. The analysis yielded a main effect on condition, $F(2,49)=3.36$, $p=.043$, partial $\eta^2=.121$. The ramp pretest was not a significant covariate, $F(1,49)=1.38$, $p=.247$. Post hoc tests revealed that the direct+reify condition constructed significantly more unconfounded experiments in the delayed posttest as compared to the direct-no reify condition (M=1.34, SE=0.53, $p=.014$, 95% CI=[0.29, 2.40]). No significant differences between direct+reify and discovery, $p=.131$, and direct-no reify and discovery, $p=.306$ were found. As shown in Figure 2, the direct+reify condition maintained its higher performance at the immediate and delayed posttest. Though the direct-no reify condition improved at the immediate posttest compared to the pretest, the improvement is lost 6 months later at the delayed posttest. Also, the discovery condition's skills on this CVS authentic inquiry task increased as time progressed.

We also compared the number of students proficient at CVS (those who scored at least 3 out of 4 on the ramp) in each condition at the delayed posttest. We used only those who did not score a perfect 4 out of 4 on the ramp pretest (49 participants). Of those remaining, 59% (10 out of 17) in the direct+reify condition, 24% (4 out of 17) in the direct-no reify condition, and 33% (5 out of 15) in the discovery condition were proficient. Though the tallies favor direct+reify over the other conditions, the differences were not significant, $\chi^2(2)=4.43$, $p=.109$. We note that in the delayed posttest, there were more than twice as many CVS-proficient students in the direct+reify condition as compared to the direct-no reify condition.

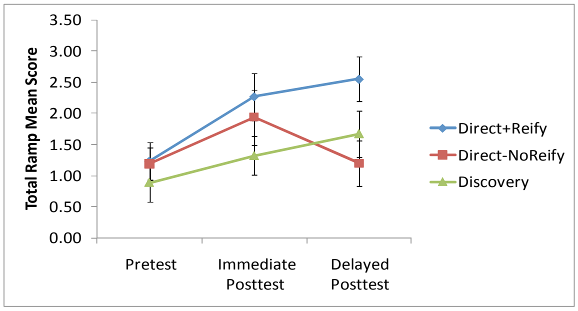

Figure 2. Means and standard errors for total ramp score by condition. The maximum score on this measure is 4.

Comparison of Ramp Items by Level of Complexity

As in our previous 2009 study, we examined performance changes over time for ramp items of differing complexity, namely singly and multiply confounded initial setups. To determine time and condition effects we conducted a repeated measures MANCOVA using singly and multiply confounded ramp items at each time point as within subjects factors, condition as a between subjects factor, and singly confounded ramp pretest score as a covariate. This design mimics the design used in our 2009 analysis. There were neither main effects on the dependent variate for time ($p=.646$) nor for the time and condition interaction, $p = .31$. However, there was a significant between subjects effect on the dependent variate for condition, Wilks $\lambda=.73$, $F(4,96)=4.18$, $p=.004$, partial $\eta^2=.148$.

ANCOVAs on each dependent measure revealed a main effect for condition for multiply confounded ramp items, $F(2,49)=7.58$, $p=.001$, partial $\eta^2=.236$ but not for singly confounded ramp items, $F(2,49)=1.18$, $p=.156$, partial $\eta^2=.073$. Post hoc comparisons between conditions for multiply confounded items revealed that the direct+reify condition created significantly more unconfounded experiments starting from a multiply confounded state over both the discovery ($M=.62$, $SE=.19$, $p=.002$, 95% $CI=[0.25, 1.00]$) and direct-no reify conditions ($M=.64$, $SE=.20$, $p=.002$, 95% $CI=[0.24, 1.03]$). Again, no significant differences were found between direct-no reify and discovery, $p=.949$. As shown in Figures 3 and 4, the direct+reify condition maintained its performance edge over time as compared to the other two conditions. Also, though the direct+reify and direct-no reify conditions were similar at immediate posttest, the direct-no reify condition did not show long-term skill in constructing unconfounded experiments as shown by the drop in performance at the delayed posttest. Again, the discovery condition continued to increase in average performance over time.

Group differences for the time-2 delayed posttest were also analyzed using a MANCOVA with singly confounded ramp pretest items as a covariate. The dependent variate comprised of singly and multiply confounded ramp items at time 2 was marginally affected by condition, Wilks $\lambda=.84$, $F(4,96)=2.21$, $p=.074$, partial $\eta^2=.084$. The covariate was also marginally significant, Wilks $\lambda=.91$, $F(2,48)=2.41$, $p=.100$, partial $\eta^2=.091$. Observing the differences more closely with ANCOVAs on each dependent measure revealed a significant effect for condition for multiply confounded ramp items, $F(2,49)=4.23$, $p=.020$, partial $\eta^2=.147$ and a marginally significant effect for condition for singly confounded ramp items, $F(2,49)=2.82$, $p=.069$, partial $\eta^2=.103$. Pairwise comparisons revealed a significant difference in constructing experiments that followed CVS starting from a multiply confounded state favoring direct+reify over direct-no reify ($M=.73$, $SE=.26$, $p=.007$, 95% $CI=[-0.21, 1.25]$), and a marginally significant difference favoring direct+reify over discovery ($M=.48$, $SE=.25$, $p=.061$, 95% $CI=[-.02, 0.97]$). The direct-no reify condition was not significantly different than discovery learning condition, $p=.350$.

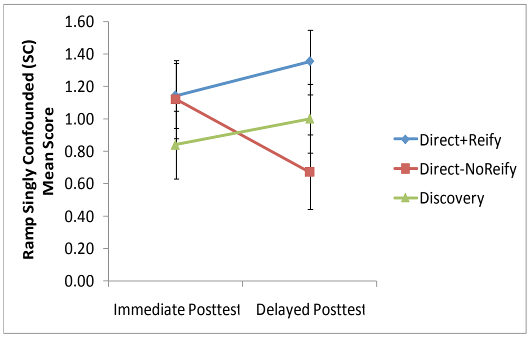

Figure 3. Means and standard errors for singly confounded ramp items (max=2). The pretest was omitted since it had only 1 singly confounded item.

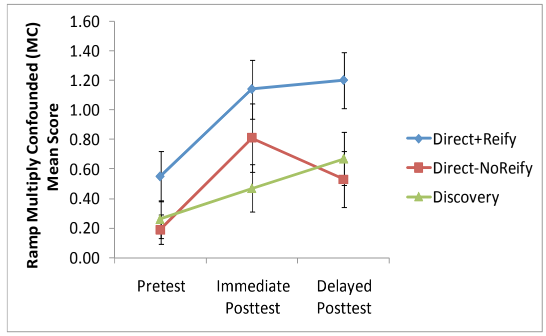

Figure 4. Means and standard errors for multiply confounded ramp items (max=2).

Discussion and Conclusions

The primary goal of this research was to compare three pedagogical approaches in terms of their efficacy at fostering robust understanding of CVS as measured by successful application of CVS and long-term retention of this skill. Specifically, we examined if direct instruction coupled with self-explanation (direct+reify) afforded any long-term benefits on CVS acquisition over direct instruction without self-explanation (direct-no reify) and discovery learning. In Sao Pedro, et al. (2009), we found that students in each learning condition did not perform significantly better than each other on CVS problem solving and overall construction of unconfounded experiments. However, those in either direct instruction condition were better at correcting the more difficult multiply confounded setups to correctly contrast variables than the discovery condition at the immediate posttest. When incorporating the results of the delayed posttest administered six months later, direct instruction with explanation yieded a deeper understanding of CVS, collapsing over time. Looking specifically at the delayed posttest, those in the direct+reify condition significantly outperformed the direct-no reify condition on authentic inquiry items starting from a multiply confounded state. Additionally, those in the direct+reify condition outperformed those in the discovery condition, but these results were not strong enough to reach statistical significance. The instruction method did not impact long-term skill to articulate CVS nor CVS problem solving skill. The differences between our direct and discovery conditions are not as drastic as those of Klahr and Nigam (2004); this may be attributed to our choice to use text as the means to communicate instruction as opposed to audio as Klahr and Nigam did.

Our results are comparable to both Strand-Cary and Klahr (2008) and Dean Jr. and Kuhn (2006). Like us, Strand-Cary and Klahr reported that in a delayed posttest, students in their “discovery” condition, on average, rose to the levels of their explicit instruction condition at constructing unconfounded experiments. We concur with their interpretation that students’ engagement at constructing unconfounded experiments, though initially unsuccessful, may have made them more aware of the necessity to control for variables. Furthermore, it may be possible that students’ initial failure primed them for future encounters with the concept (Kapur, 2008). Further research needs to be conducted to empirically test these claims. Nonetheless, our collective findings suggest that it is possible for unguided exploration to reach levels comparable to explicit instruction with self-explanation, albeit more slowly. Additionally, Dean Jr. and Kuhn (2006) found that participants who practiced CVS regularly, irrespective of whether they received direct instruction or not, were better able to apply the strategy and make inferences when compared to those receiving direct instruction and not practicing. In their study, the practice condition dyads provided explanations to each other and entered them into the computer; these activities could have significantly affected students’ knowledge of CVS. In other words, it is difficult to ascertain whether the robustness of their findings is due to practicing the skill or providing the explanation.

In this paper, we have shown that the reification process (self-explanation), when added to direct instruction of the skill, supports both efficient acquisition and long-term retention of CVS, as measured by authentic inquiry performance. Of interest, we also found that on the same measurement, a nontrivial number of students in our unguided discovery condition were also able to perform, on average, at levels comparable to the direct+reify condition 6 months after the initial intervention. To address which behaviors within each intervention led to robust acquisition of CVS, we are looking at the quality of self-explanations, student responses, and all log files of student interactions within the ramp microworld. This will allow us to characterize the affordances of each condition for learning CVS and examine interactions with student variables such as prior knowledge of inquiry, in particular, knowledge of designing and conducting experiments. A fuller description such as this will provide important data towards our eventual goal of developing adaptive scaffolding support for CVS as it is a necessary but not sufficient skill for conducting inquiry (Gobert et al., 2007; Gobert et al., 2009).

References

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.

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

Chi, M., & VanLehn, K. (2007). Accelerated Future Learning via Explicit Instruction of a Problem Solving Strategy. *13th International Conference on Artificial Intelligence in Education* (pp. 409-416). Amsterdam: IOS Press.

de Jong, T. (2006). Computer Simulations - Technological advances in inquiry learning. *Science, 312* , 532-533.

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

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.

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.

Kapur, M. (2008). Productive Failure. *Cognition and Instruction, 26(3)* , 379-424.

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., & Dunbar, K. (1988). Dual search space during scientific reasoning. *Cognitive Science, 12*, 1-48.

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

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Society for Research in Child Development Monographs 60(4, Serial No. 245)* .

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

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

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.

Schunn, C., Raghavan, K., & Cho, N. (2007, April 9-13). Domain-general Learning Accelerators in Middle-School Science. Chicago, IL: Presented at the annual meeting of the American Educational Research Association.

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

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., et al. (2005). The Andes physics tutoring system: Lessons Learned. *International Journal of Artificial Intelligence and Education, 15(3)*, 1-47.

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.