

Comparing Pedagogical Approaches for Teaching the Control of Variables Strategy

Michael A. Sao Pedro¹ (mikesp@wpi.edu)

Janice D. Gobert^{2,1} (jgobert@wpi.edu)

Neil T. Heffernan¹ (nth@wpi.edu)

Joseph E. Beck¹ (josephbeck@wpi.edu)

¹Department of Computer Science and ²Department of Social Sciences and Policy Studies
Worcester Polytechnic Institute
100 Institute Rd. Worcester, MA 01609 USA

Abstract

In this study, an extension of Klahr and Nigam (2004), we tested 177 middle school students' on their acquisition of the control of variables strategy (CVS) using an interactive virtual ramp environment. We compared the effectiveness of three pedagogical approaches, namely, direct instruction with reification, direct instruction without reification, and discovery learning, all of which were authored using the ASSISTment system. MANCOVAs showed that all conditions performed equally on a CVS multiple-choice post test, but that the 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.

Keywords: scientific inquiry learning; web-based interactive environment; learning with microworlds; direct vs. discovery learning; control of variables strategy.

Introduction

Reform efforts in education are consistent with the National Science Education Standards (1996) about the importance of inquiry skills, namely identifying questions that guide scientific investigations, using technology to improve investigations, formulating and revising scientific explanations and models using logic and evidence, recognizing and analyzing alternative explanations, and communicating and defending a scientific argument. Although there is currently a large body of research on inquiry learning, there is mixed results regarding efficacy. For example, if inquiry tasks are too open-ended, students often become lost and frustrated and their confusion can lead to misconceptions (Brown & Campione, 1994). As a result, teachers spend considerable time scaffolding students' procedural skills (Aulls, 2002) making it difficult to tailor to individual students in real time within a classroom setting (Fadel, Honey, & Pasnich, 2007). Furthermore, during inquiry learning students can have many false starts (Schauble, 1990) and may have difficulties designing effective experiments, forming testable hypotheses, translating theoretical variables from their hypotheses into variables, adequately monitoring what they do (de Jong, 2006), linking hypotheses and data, and drawing correct conclusions (Klahr & Dunbar, 1988; Kuhn, Kuhn, Garcia-Mila, Zohar, & Andersen, 1995). Ultimately, these difficulties can impede the learning of targeted scientific principles (Kirschner, Sweller, & Clark, 2006).

On the other hand, some researchers have successfully shown that students can effectively learn scientific inquiry strategies. For example, Klahr and Nigam (2004) focused on the control of variables strategy (CVS), a procedure for setting up experimental contrasts such that only one variable is tested and all others are held constant. In Klahr and Nigam's study, third and fourth grade students participated in one of two learning conditions. In the discovery condition, students used an actual ramp apparatus and were asked to construct unconfounded experiments to determine which factors led to making a ball roll further down a ramp. In the direct condition, the experimenter set up several experiments and asked students to determine if an experiment was confounded while thinking aloud. The authors found that students who were taught CVS in the direct learning condition significantly outperformed those in the discovery learning condition on a near-transfer test for CVS. These results suggest that the purported benefits of discovery learning, particularly the deeper learning, may not always occur. The authors used their data to suggest a mechanism called "path independent learning", meaning that if learning occurs, performance will be the same irrespective of the teaching method.

The Klahr and Nigam (2004) findings, though, are not without critique. For example, recent studies have been conducted in which authors unpack the affordances of inquiry-oriented forms of learning (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 or what goals the strategy accomplishes.

In this paper, we seek to formally compare discovery versus direct learning of the CVS using a learning environment in which we can rigorously control for all variables in all the learning conditions. Furthermore, since our technology allows us to randomly assign students to condition within the same class, our findings are not subject to criticisms concerning teacher variables and classroom culture variables. This represents progress in the debate about discovery versus direct learning because in many studies in which these two forms of pedagogy are contrasted, there are extraneous variables that also differ across the two conditions.

Our research aims to recreate and extend Klahr and Nigam's (2004) CVS conditions, but we provide instruction using a virtual environment. Similar to us, Triona and Klahr (2003) compared the effects of physical versus virtual

materials on elementary school students' skills at designing controlled experiments and found no significant performance differences. They controlled for instruction by using a human teacher in both conditions. In our study, we embody the direct and discovery conditions entirely within the learning environment. To do this, we constructed a virtual ramp environment by extending the ASSISTment system currently used for Math (Gobert, Heffernan, Ruiz, & Kim, 2008)

Our study has several goals. First, we aim to compare direct and discovery conditions entirely using a virtual learning environment. We also extend the original experiment (Klahr & Nigam, 2004) by adding a third condition, namely direct no-reify, which removes prompting for student explanations, since self-explanation could have played a role in students' acquisition of CVS (Chi, 1996) in Klahr and Nigam's study. This enables us to empirically test if "thinking aloud" affects the acquisition of CVS in our study.

Method

Participants

Participants were 97 seventh- and 80 eighth-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). At this school, each grade level had its own science teacher who taught five different class sections. All sections participated in the experiment.

Materials

Our study makes use of the ASSISTment system, a web-based intelligent tutoring system for mathematics that simultaneously assesses students while they receive assistance with particular math problems (Heffernan, Turner, Lourenco, Macasek, Nuzzo-Jones, & Koedinger, 2006). It provides the capability to easily create static content, i.e. multiple choice, fill-in-the-blanks, and open response questions with embedded images, videos, and Flash files. It also tracks student information like skill progression for math, time taken per question, percentage of correct answers, and number of hints received, and provides summary reports teachers can use to evaluate their students.

The ramp environment, shown in Figure 1, was developed using the OpenLaszlo framework (www.openlaszlo.org). We create different kinds of questions by embedding the ramp environment within an ASSISTment problem. 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 the balls to roll down the ramp. Depending on each ramp's settings, the balls will roll different distances down the ramp. The "submit" button allows participants to submit their final answer. Pressing

"reset" causes the balls to be placed back on the ramp and clears the distance rolled. Each time the "run" or "submit" button is pressed, 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.

Students interact with our ramp environment differently than the physical materials in Klahr and Nigam (2004) and Triona and Klahr (2003). In our environment, variable values are changed using combo boxes rather than adding and removing ramp pieces. More importantly, in the other studies participants built ramp configurations starting from a blank slate. In our environment, a configuration is always prebuilt. In other words, there is always an initial condition students must change in order 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), and/or *uncontrasted* (the target variable is unchanged).

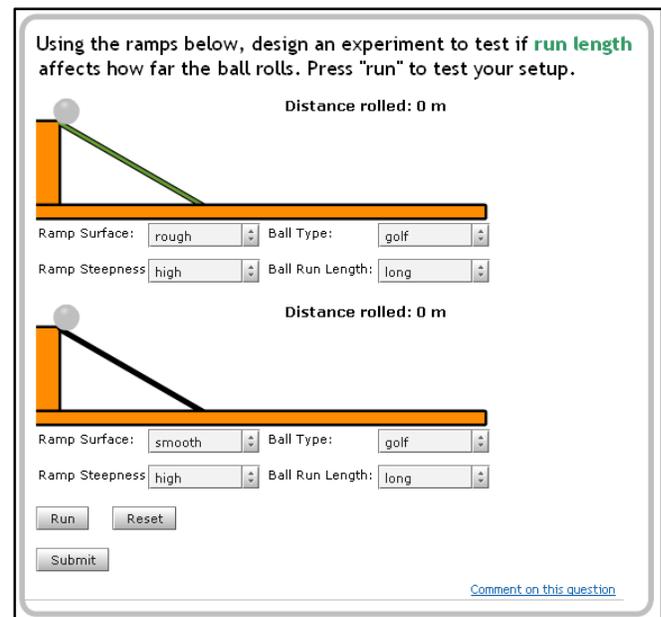


Figure 1: Typical ASSISTment question using the ramp environment. The initial setup shown here is uncontrasted and confounded because target variable, run length, is the same for each ramp and ramp surface is not controlled.

To emulate direct instruction with and without reification and discovery learning, we designed a series of ASSISTment questions that use the interactive incline plane environment. 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. We describe these questions in more detail in our procedure.

Procedure

The experiment took place over three classroom periods on three non-consecutive days. Each period was about 60 minutes long. Day 2 took place 2 weeks after Day 1. Day 3 took place the day after Day 2. Students in each section grade were randomly assigned by the ASSISTment system to either the direct+reify, direct-no reify, or the discovery condition. Students worked in the computer lab during their regular class time, one student per computer. Most students had familiarity with the ASSISTment system from math class, but none had previously used the ASSISTment system for science nor had they used our interactive ramp environment.

Day 1 consisted of two phases: inquiry pre-assessment and survey. In the inquiry pre-assessment phase participants answered standardized-test style multiple choice and open-response questions. Three multiple-choice questions pertained directly to CVS.

Day 2 consisted of several phases: introduction to the ramp environment and its variables, ramp pretest, intervention (direct+reify, direct-no reify, discovery), ramp posttest, and inquiry post-test. During the ramp pretest phase, all participants, regardless of condition, demonstrated their understanding of CVS by attempting to construct four unconfounded setups to test if a particular variable affected how far the ball rolled. Participants could modify and run their experiment as many times as desired. Once satisfied, they submitted their setup as their final answer. Each initial ramp setup began with a different initial configuration that focused on different target variables. The first was unconfounded, the third singly confounded, and the other two multiply confounded. Additionally, the first two questions focused on the steepness variable and the second two focused on the run length variable. For each item, the target variable was set to be uncontrasted thus requiring students to change this setting. No feedback was given to participants during the introduction and ramp pretest phases.

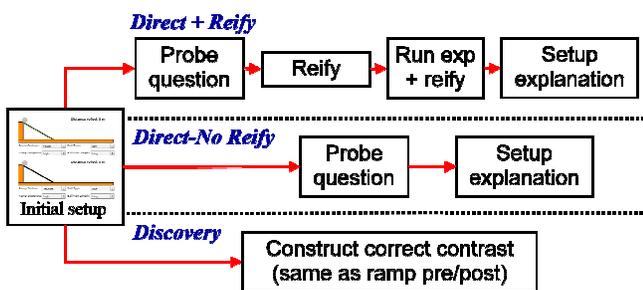


Figure 2: Day 2 experimental design. Both direct conditions received an initial CVS primer before answering questions.

During the intervention phase, students practiced CVS with six new ramp configurations. The same six same ramp setups were presented in the same order, some confounded, some unconfounded, irrespective of condition. All setups focused on run length and steepness. The nature of the interaction with the ramp environment changed depending

on the condition as shown in Figure 2. In the direct+reify condition, students were first asked to read an overview of CVS with good and bad examples of ramp setups. Next, participants evaluated ramp setups had correctly controlled for variables or not. For each setup, participants responded if they could “tell for sure” if a variable affected how far a ball would roll. Students could see both ramps and each variable’s values but could not run the experiment nor change the variable values; they simply responded “yes” or “no”. Next, they answered a free response question asking them to justify their choice. For the same ramp setup, students then were allowed to run the experiment as many times as desired and explained again if they could tell for sure that target variable affected how far the ball would roll after running the experiment. Finally, the students read an explanation why the experiment was confounded or unconfounded for the target variable. If the setup was confounded, students were told exactly 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 given the same six ramp setups as the direct+reify and direct-no reify conditions. Like the pretest, they were instructed to create experiments that tested if a particular variable affected how far the ball rolled. Again, they could run the experiment and change variable values as many times as desired until satisfied. The discovery condition students were not given the initial CVS explanation nor were they given any feedback about the correctness of their experimental setups.

Conventionally, “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, whereas discovery embodies what would be referred to as unguided inquiry.

During the ramp posttest phase, all participants attempted to construct unconfounded experiments for three target variables, two of which had not yet been the focus: run length, surface, and ball type. No feedback was given to the students on the correctness of the experimental setups.

During the inquiry posttest phase, students were asked the same three CVS questions as on Day 1 in addition to open-response questions about general CVS strategy. Students who did not complete all the activities on Day 2 were allowed to continue working on the next day (Day 3).

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 scored 1, 0 otherwise. In our analysis, we considered only the three CVS multiple choice pre/post test questions and four pre/post test ramp setups; the open response items are currently being scored and thus are not a part of this report.

Table 1: Summary of student performance on pretest and posttest items.

	Max	<i>Direct+Reify</i> (N=45)		<i>Direct-No Reify</i> (N=42)		<i>Discovery</i> (N=43)	
		<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Inquiry Pretest*	3	1.73	0.99	1.37	1.00	1.43	0.82
Ramp Pretest	4	1.49	1.42	1.31	1.46	1.02	1.44
Unconfounded	1	0.47	0.51	0.38	0.49	0.37	0.49
Single confound	1	0.40	0.50	0.45	0.50	0.33	0.47
Multiple confounds	2	0.63	0.82	0.48	0.74	0.33	0.64
Inquiry Posttest	3	1.98	0.94	1.50	0.97	1.67	0.97
Ramp Posttest	4	2.28	1.80	2.12	1.71	1.40	1.45
Unconfounded	0	-	-	-	-	-	-
Single confound	2	1.12	0.93	1.10	0.91	0.88	0.91
Multiple confounds	2	1.16	0.95	1.02	0.90	0.51	0.70

* The mean and standard deviation computations for inquiry pretest include imputed missing values.

Results

Our analyses focus on which condition(s) yielded better performance on the inquiry and ramp posttests. In particular, we address whether there were differences between: 1) our direct+reify vs. direct-no reify conditions and 2) direct-no reify vs. discovery learning conditions. We discarded students on individual education programs (IEPs) and one additional student who used an incorrect web browser leaving 133 participants, 74 seventh grade and 59 eighth grade students. Eight students were classified as English language learners (ELLs). Due to absences, time constraints, and students accidentally skipping questions, 17 students (5 direct+reify, 3 direct-no reify, and 9 discovery) did not complete the inquiry pretest on Day 1. For similar reasons, 3 students did not complete the ramp pretest, 5 did not complete the ramp posttest, and 4 did not complete the inquiry posttest on Days 2 and 3. Due to the large number of students missing inquiry pretest data, we imputed missing values using the mean. Although our experiment ran in a total of 10 different sections, we did not model section in any of our analyses because all students within a grade had the same science teacher, sections were heterogeneous, and students were randomly assigned to condition by ASSISTment within each class. Means and standard deviations for the conditions on both the pre- and posttests are shown in Table 1.

To determine if participants in three conditions were significantly different at pretest, a one-way between-subjects multivariate analysis of variance (MANOVA) was conducted using inquiry pretest and ramp pretest scores as dependent measures and condition as a factor. There were no significant differences at pretest across the conditions, Wilks $\lambda=.96$, $F(4,252)=1.36$, $p=.247$, partial $\eta^2=.021$. We also analyzed if students learned overall amongst the conditions by comparing aggregated pretest performance and aggregated posttest performance. There was a reliable gain ($M = .86$), slightly less than one question, on the 7-item tests, $t(127)=5.07$, $p<.001$, 95% $CI=[.53, 1.20]$.

Testing for overall posttest differences

To determine which learning condition led to improved CVS inquiry multiple choice and ramp posttest scores we performed a one-way between-subjects multivariate analysis of covariance (MANCOVA) using the inquiry posttest and ramp posttest as dependent measures and learning condition as a factor. We considered grade level, ELL status, inquiry pretest, intervention time, and ramp pretest as potential covariates, but only the ramp pretest was used as a covariate thereby holding its effects on the post-test constant. We chose it because it was the only variable that significantly correlated with the inquiry post test, $r=.36$, $p<.001$ and the ramp posttest, $r=.58$, $p<.001$. Intervention time was not used as a covariate because it did not correlate significantly with the inquiry posttest, $r=.14$, $p=.114$, or the ramp posttest, $r=.17$, $p=.059$. The dependent variate was significantly affected by condition, Wilks $\lambda=.92$, $F(4,246)=2.54$, $p=.041$, partial $\eta^2=.040$. Follow-up univariate ANCOVAs, shown in Table 2, revealed that the effects of condition were not significant for the ramp posttest, $F(2,124)=2.32$, $p=.102$, partial $\eta^2=.036$, and marginally significant for the inquiry posttest, $F(2,124)=2.49$, $p=.087$, partial $\eta^2=.039$.

In line with Klahr and Nigam's (2004) analysis, we also determined if the number of ramp CVS masters, those who scored at least 3 out of 4 on the ramp posttest, differed between groups using only participants who did not score a perfect 4 out of 4 on the ramp pretest (114 participants). Of those remaining, 19 out of 36 in the direct+reify condition, 15 out of 37 in the direct-no reify condition, and 7 out of 39 in the discovery condition were deemed CVS masters. The differences between these groups was significant, $\chi^2(2)=10.15$, $p=.006$. Post hoc tests revealed that there were significantly more masters in the direct+reify condition as compared to discovery, $\chi^2(1, N=75)=10.03$, $p=.002$ and the direct-no reify condition as compared to discovery, $\chi^2(1, N=76)=4.71$, $p=.030$. However, there was no significant difference between the and direct-no reify conditions, $\chi^2(1, N=73)=1.10$, $p=.295$.

Table 2: Univariate ANCOVA factor and covariate significance for inquiry and ramp posttest items.

Source	Posttest	df	F ^a	p	part. η ²
Ramp Pre	Inq	1	18.63	<.001	.131
	Ramp	1	59.50	<.001	.324
Condition	Inq	2	2.49	.087	.039
	Ramp	2	2.32	.102	.036

a: *df error*=124. *MSE Inquiry*=0.81, *MSE Ramp*=1.88.

Analysis of ramp items by level of complexity

We also investigated if certain posttest items yielded different learning gains across conditions. In particular, we tested whether performance was different for items involving singly confounded or multiply confounded ramp items, the latter of which are more complex, using a MANCOVA. Singly confounded (SC) and multiply confounded (MC) ramp posttest items were dependent variables and condition was the factor. ELL status, grade, ramp pretest items for unconfounded (UC), singly confounded (SC), and multiply confounded (MC) were considered as covariates. Only the singly confounded pretest item score was used as a covariate because out of the three ramp pretest scores that significantly correlated with each other ($.45 \leq r \leq .64$, $p < .001$), it correlated the highest with the ramp SC posttest ($r = .51$, $p < .001$) and ramp MC posttest ($r = .56$, $p < .001$) scores. Again, intervention time was not used as a covariate because it did not correlate significantly with the SC ramp posttest, $r = .11$, $p = .213$, or MC ramp post test, $r = .21$, $p = .018$. Condition was again significant with respect to the combined dependents, Wilks $\lambda = .863$, $F(4,246) = 4.70$, $p = .001$, partial $\eta^2 = .071$. ANCOVAs were computed for each ramp posttest dependent and the results are presented in Table 3. For the singly confounded items, i.e., the less complex items, condition was not significant, $F(2,124) = 0.30$, $p = .622$. However, condition was significant for the more complex (multiply confounded) items, $F(2,124) = 3.76$, $p = .001$, partial $\eta^2 = .107$.

Main effects comparisons with Bonferroni correction on confidence intervals yielded the following results. First, both the direct+reify condition, $M = 0.58$, $SE = 0.15$, $p = .001$, 95% $CI = [0.21, 0.96]$ and direct-no reify condition, $M = .39$, $SE = 0.16$, $p = .041$, 95% $CI = [0.01, 0.76]$ significantly outperformed the discovery condition. Second, no significant differences were found between the direct+reify and direct-no reify conditions, $M = 0.20$, $SE = 0.15$, $p = .624$, 95% $CI = [-0.18, 0.57]$.

Discussion and Conclusions

Our principle goal in this study was to develop a learning environment to teach CVS and empirically test its efficacy in three different learning conditions: direct+reify, direct-no reify, and discovery. It is important to note that in this study

Table 3: Univariate ANCOVA factor and covariate significance for singly confounded (SC) and multiply confounded (MC) ramp posttest items.

Source	Ramp posttest	df	F ^a	p	part. η ²
Pre SC	Ramp MC	1	57.80	<.001	.318
	Ramp SC	1	44.56	<.001	.264
Condition	Ramp MC	2	7.46	.001	.107
	Ramp SC	2	0.48	.622	.008

a: *df error*=124. *MSE RampMC*=0.50, *MSE RampSC*=0.62

we are prioritizing the inquiry skill for control for variables, which is a skill under the “designing and conducting scientific investigations” strand, over other skills listed by the National Science Education Standards (1996). Specifically, we argue that the control for variables strategy provides a good conceptual and procedural anchor for learning inquiry processes and skills. It is critical that students understand the importance of controlling for variables when doing experiments, and thus important that CVS be explicitly taught, particularly because it has been shown that scientifically naïve adults do not understand control for variables (Kuhn, 2005).

In this study, we have shown that a virtual environment can be used to teach the control for variables strategy. Although differences between the three learning conditions were not statistically different on the overall inquiry and ramp post-tests, significant effects were found favoring both direct conditions (direct-no reify and direct+reify) over the discovery condition for more complex ramp posttest questions. It is important to note that effects due to prior knowledge were controlled for by using the ramp pre-test as a covariate; thus, differences are not due to pre-existing differences in students. Thus, direct instruction is a more beneficial method for teaching CVS when compared to discovery learning. Our results are compatible with Klahr and Nigam’s in that we showed learning gains for those in both direct conditions over the discovery condition for more complex items involving multiply confounded items. We also found that each direct variant produced more CVS experts than the discovery condition, in accordance with Klahr and Nigam’s findings.

Our study also extended that of Klahr and Nigam by adding a third learning condition, namely direct-no reify. By adding this condition, we were able to compare these data to the direct+reify conditions and thus address whether having students’ reify their understanding of CVS in the form of explanations would yield significantly better performance on posttest understanding of CVS. No statistically significant differences were found between the two direct learning conditions, meaning that for ramp CVS items, generating explanations in the direct+reify condition did not provide an added benefit over direct-no reify. We will analyze the open-ended responses from the inquiry post-test in the two direct conditions to evaluate whether students in the direct+reify condition provided qualitatively better

explanations over those in the direct-no reify conditions. If borne out, these results could provide insight into the role of explanation in the acquisition of inquiry skills and CVS in particular. These data will also provide evidence sought by Kuhn (2005) about the richness of students' understanding of CVS. Furthermore, we argue that it is likely that explanation tasks in which students reify their learning, referred to as communication in the NSES inquiry strands, may be beneficial for other inquiry strands. Specifically, inquiry processes are both complex and multi-faceted; it is our belief that students need to be appropriately scaffolded and have competency honed using a progressive skill building approach. The degree of scaffolding needed, the transfer of inquiry skills to new domains, and interactions between inquiry skills and student characteristics are all empirical issues addressed in our program of work.

Our results also support Triona and Klahr's (2003) findings that CVS can be acquired by engaging with virtual materials. We extended this work by embedding variants of the authors' step-by-step tutorials of CVS within the computer. As they point out, several design decisions may influence the learning outcomes such as using text versus audio and how or if open-ended responses are necessary for learning CVS. Our effects between 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 previously stated, there currently is a debate in the science education community, which has juxtaposed direct instruction with open-ended discovery. Under the umbrella of open-ended discovery, all forms of inquiry (Kirschner, Sweller, & Clark, 2006) are, we feel, erroneously grouped together as open-ended constructivist approaches. We recognize that inquiry approaches range from more to less open-ended activities in terms of task structure, and claim that scaffolding science inquiry processes, strategies, etc., for learning science is *not* equivalent to direct instruction since we believe that this can both foster the skills themselves and well as be generative in terms of supporting science content (Gobert et al., 2008).

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

Aulls, M. (2002). The Contributions of Co-Occurring Forms of Classroom Discourse and academic Activities to Curriculum Events and Instruction. *Journal of Educational Psychology, 94*(3), 520-538.

Brown, A., & Campione, J. (1994). Guided discovery in a community of learners. In K. M. (Ed.), *Classroom lessons: integrating cognitive theory and classroom practice*. Cambridge, Massachusetts: MIT Press.

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

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

Fadel, C., Honey, M., & Pasnick, S. (2007). Assessment in the Age of Innovation. *Education Week, Volume 26* (38), 34-40.

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

Heffernan, N., Turner, T., Lourenco, A., Macasek, M., Nuzzo-Jones, G., & Koedinger, K. (2006). The ASSISTment builder: Towards an analysis of cost effectiveness of ITS creation. *Proceedings of the 19th International FLAIRS Conference*, (pp. 515-520). Melbourne Beach, Florida, USA.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, 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., & 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). *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.

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).

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

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. (1997, April 9-13). Domain-general Learning Accelerators in Middle-School Science. Chicago, IL: Presented at the annual meeting of the American Educational Research Association.

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.