Supporting Middle Schoolers' Use of Inquiry Strategies For Discovering Multivariate Relations In Interactive Physics Simulations

Luke D. Conlin, Nicole Hallinen, Daniel L. Schwartz, Stanford University, Graduate School of Education, Wallenberg Hall

Email: lconlin@stanford.edu, hallinen@stanford.edu, danls@stanford.edu

Abstract: Within research on students' inquiry into related variation, several researchers have pointed out the importance of students understanding multiple variable relations. So far, the Control of Variables Strategy (CVS) has demonstrated only limited success in supporting students' discovery of multiple variable relationships. In this report, we present an alternate strategy, which we call the General Principle Strategy (GPS). We report on preliminary results of a classroom study where we taught students in two conditions to use CVS or GPS, respectively, in the context of several physics topics. We find evidence that both strategies help students figure out the multivariable relationship underlying the working of a balance scale, as inferred from associations between their performance on a written posttest and on a computer game-based posttest. Based on these results, GPS shows promise as an effective way of teaching multiple variable relations that underlie a wide variety of physics phenomena.

Introduction

Engaging in scientific inquiry allows students to learn science content while participating in the epistemic practices of science. diSessa (2008) has identified two distinct but complementary modes of inquiry prevalent in the literature, which tap into different aspects of authentic disciplinary practices: inquiry into the meaning of concepts, and inquiry into related variation. The latter involves empirically discovering relations between variables, such as the relation between the range of a projectile and its initial speed.

One line of research on inquiry in science classrooms has identified specific strategies that scientists use to figure out the causal relations between variables and has explored the effectiveness of explicitly teaching these strategies to support students' inquiry (Chen & Klahr, 1999; Ford, 2005; Kuhn, Pease, & Wirkala, 2009). The most prominent inquiry strategy employed in these studies has been the *control of variables strategy* (CVS). CVS involves discovering relations between variables by designing controlled experiments, changing only one variable at a time to make unconfounded comparisons.

Many studies have reported success in teaching CVS to learners in a variety of age groups, who learn to set up unconfounded comparisons and to draw correct inferences from them (Chen & Klahr, 1999; Ford, 2005; Kuhn et al., 2009). These studies have primarily demonstrated the usefulness of CVS for discovering single variable relationships, where only the main effects of a variable are considered, and not its interaction with other variables. For instance, students might use the CVS strategy to find out whether the length of a spring, its width, or how much weight is hung from it affects how far a spring stretches, but do not explore whether the effect of hanging a weight changes based on, say, the width of the spring (Chen & Klahr, 1999; Ford, 2005). Many relations in science involve multiple interacting variables, and as Kuhn (2007) has pointed out, CVS may not be sufficient for unpacking these relationships. So far, there is little evidence that students' learning of CVS helps with their discovery of the relationship between multiple interacting variables (Kuhn et al., 2009; Kuhn, 2007), although this could be due to a lack of instructional supports for extending CVS to handle multiple interacting variables.

Much of the research on students' inquiry into related variation has focused on the hypothetico-deductive approach to science, which is the logic of inference underlying CVS. This approach begins with the formulation of a hypothesis that is then used to deduce observational consequences. Much less work has been done to explore students' use of an equally valid logic of inference, which has played a comparably important role in science: the inductive approach (Shemwell, Chase, & Schwartz, under review). Induction begins with making observations and synthesizing an underlying principle or explanation. While science educators generally recognize the importance of inductively searching for patterns in data, little work has been done to investigate how to support students in conducting such a search in a systematic way.

In this paper, we present an inductive strategy, *general principle strategy* (GPS), which shows promise for supporting students' inquiry into related variation, particularly for discovering the relationships between multiple variables. This strategy has roots in the history and philosophy of science, dating at least back to Bacon (Shemwell, Chase, & Schwartz, under review), and has strong connections to modern accounts of unification and coherence-seeking in science. The general approach involves examining all the data to find one underlying general explanation. This can apply to a broad array of contexts where CVS may be impracticable (e.g., the historical discovery that the evening star and the morning star were in fact the same object, Venus), but it can also be used as an alternative for making comparisons that establish relations between variables.

GPS offers another way of making unconfounded comparisons. Instead of making pairwise comparisons based on dependent variables (as in CVS), the GPS approach to is to make comparisons across cases based on a common outcome, then to look for common characteristics. The logic of GPS involves using the dependent variable to make inferences about the independent variables, while for CVS the logic of inference proceeds from independent variables to the dependent. For instance, given the top speeds of a set of airplanes with different wing lengths, body shapes, and tail configurations, the GPS approach would be to look at the fastest planes and see what their common characteristics are. The CVS approach would be to pick a characteristic (wing length) and vary only that characteristic to see if the speeds are different.

Other studies have explored ways of supporting students in looking across cases to find a general explanation, for example, by having students invent an index that could apply to multiple contrasting cases. Schwartz, Chase, Oppezzo, & Chin (2011) compared two instructional methods for teaching 7th & 8th grade students the ratio concept underlying density. They found that students who were instructed to invent a "crowdedness" index that could apply to multiple contrasting cases better learned and applied the ratio concept to new physics topics compared with students who were told the ratio concept and given cases to practice. Chi, Dohmen, Shemwell, Chin, Chase, & Schwartz (2012) found improved learning outcomes for undergraduate students who were told to invent a general explanation that can predict the range of several contrasting cases of projectiles. Chase, Shemwell, & Schwartz (2010) compared the general explanation strategy with a Predict, Observe, and Explain (POE) strategy during 50-minute lesson using a physics simulation related to Faraday's law. They added explicit support of the general explanation strategy by providing an example from another domain (buoyancy). They found that students who were guided to seek a general explanation across the cases developed a deeper understanding of the vector component nature of magnetic flux than POE students. In these studies, the supports for students seeking a general explanation were largely embedded in the task, rather than being at the focus of extended, explicit instruction.

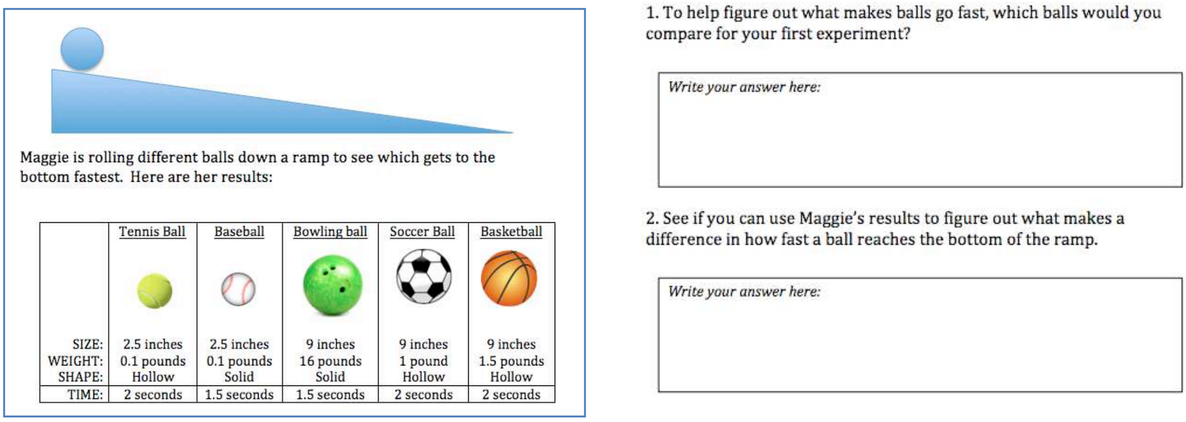
Given that several studies have shown that explicit instruction of strategies can improve students' learning and transfer of the strategies (Chen & Klahr, 1999), it is of interest to know whether explicit instruction of GPS could enhance students' learning of the strategy. In what follows, we report on the results of a study in which we taught middle school students either the CVS or GPS strategy in the context of several physics topics, over several weeks. We focus on the results of a posttest item designed to assess their use of an inquiry strategy on a novel physics topic. We report on several interesting associations between their choice of inquiry strategy on this item and their performance on a subsequent computer game-based assessment of their discovery of a multivariable relationship. These results substantiate GPS as a useful strategy for supporting students' inquiry into multivariable relationships.

Methods: Teaching and Assessment

In the present study, we taught four classes of middle school students (132 total) one of two strategies for figuring things out in science (CVS or GPS) during seven 50-minute sessions over a three-week period. Each class was randomly split into two conditions, stratified by class grade and gender. We refer to these conditions as CV or GP to disambiguate them from the strategies. The principle difference between conditions was the strategy they learned for doing inquiry. The CV condition received explicit CVS instruction applied to a variety of physics topics, including projectiles, buoyancy, and collisions, with a focus on learning the content through inquiry. The GP condition received explicit GPS instruction applied to the same sequence of physics topics with the same focus on figuring things out through inquiry. The lessons were taught by two instructors, who each taught 2 classes in each condition to counteract class and teacher effects.

In both conditions, the instruction included a variety of activities such as hands-on explorations, worksheets, and computer simulations of physics phenomena. For example, on the 5th day of instruction, students in both conditions were given simple pendulums (strings with metal washers), and asked to figure out what matters for how quickly a pendulum goes back and forth. In the CV condition, students were encouraged to pick a variable (mass, length, angle, etc.) that might affect how quickly a pendulum will go back and forth, and to test its effect by making comparisons that vary only variable at a time. In the GP condition, the students were also tasked with finding out what affects the pendulum period, but their instructions were to conduct their experiments to find multiple ways to make two pendulum swing at the same rate. Both groups were encouraged to determine the causes of changes in the pendulum period. The sequence of lessons for both groups moved from using inquiry strategies to make causal inferences about single variables to considering multiple variable relations.

After the sequence of lessons, each class took a written posttest that included an item assessing their use of either strategy (CVS or GPS) in the context of a new physics topic, racing different balls down a ramp (see Figure 1). The item presented data on five balls that were rolled down the ramp, including their size, weight, shape, and the outcome (how long it took to them to roll down the ramp). The item asked the students to: (i) decide which balls they would compare to figure out what makes them go fast (the Ramp Comparisons task), and (ii) use the data to decide what matters for how fast a ball reaches the bottom (the Ramp Conclusions task).


Maggie is rolling different balls down a ramp to see which gets to the bottom fastest. Here are her results:

	Tennis Ball	Baseball	Bowling ball	Soccer Ball	Basketball
SIZE:	2.5 inches	2.5 inches	9 inches	9 inches	9 inches
WEIGHT:	0.1 pounds	0.1 pounds	16 pounds	1 pound	1.5 pounds
SHAPE:	Hollow	Solid	Solid	Hollow	Hollow
TIME:	2 seconds	1.5 seconds	1.5 seconds	2 seconds	2 seconds

1. To help figure out what makes balls go fast, which balls would you compare for your first experiment?

Write your answer here:

2. See if you can use Maggie's results to figure out what makes a difference in how fast a ball reaches the bottom of the ramp.

Write your answer here:

Figure 1. The "Ramp" written posttest item

The day after the written posttest, students took a posttest in the form of a computer game adapted from a physics simulation of a balance scale (Wieman, Adams, & Perkins, 2008). The game included a Challenge mode and an Exploration mode (see Figure 2). In Challenge mode, the students were presented with a sequence of eight challenges: they had to predict whether a given configuration would tip left, tip right, or balance in the middle. In Exploration mode, the students were free to place bricks anywhere on the balance scale and see what happened. Their ultimate goal was to answer eight challenge problems correctly in a row.

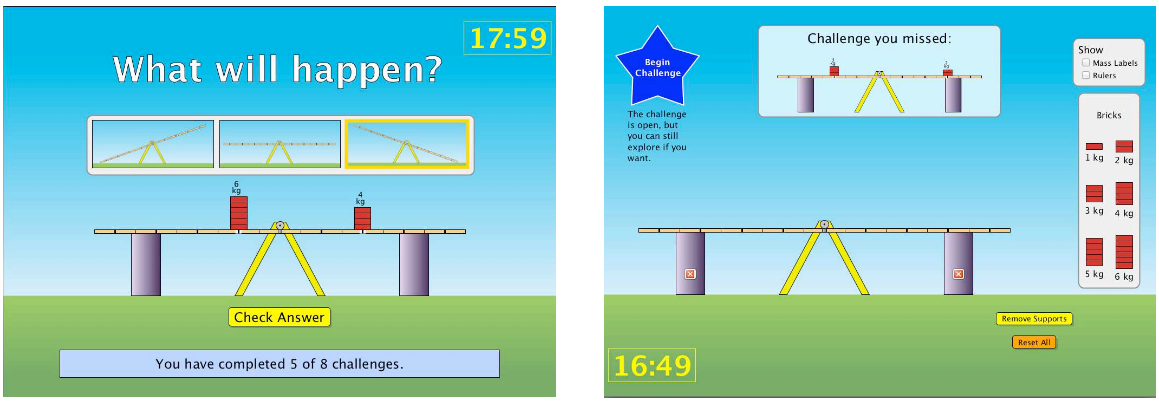


Figure 2. Balance Act posttest Challenge mode (left) and Explore mode (right)

To make predictions about whether the two sides balance, students need to consider multiple variables simultaneously, i.e., the weight on each side and their distances from the fulcrum. The sequence of challenges started off testing just the main effects of each variable (e.g., same amount of bricks on both sides, but farther out on one side) but increased in difficulty to include variable interactions (e.g., one side has more bricks but they are closer to the fulcrum, as in Figure 2). To complete the Challenge mode, the students had to make eight correct predictions in a row. As soon as they got one wrong, they were returned to Exploration mode along with a display of the configuration they missed. They were free to explore, but they could choose to re-enter Challenge mode at any time. When they returned to Challenge mode they had to start again with a whole new set of eight challenges to get through. Performing perfectly in Challenge mode is not likely without figuring out the multiplicative relationship of weight and distance, and so all told, the game serves as an assessment of students' preparedness to learn the multivariate relationship.

Data & Analysis

Of the 132 middle school students, 29 did not return consent forms and were excluded from the analysis, as were 3 students who were absent from either day of posttesting, leaving a sample of n = 100. In what follows, we present an analysis of students' performances on both the written and computer-based posttests. First we explain the coding scheme for responses to the Ramp Comparisons and the Ramp Conclusions tasks on the written posttest. Then we discuss how students' performance on the Balance Act computer-based assessment game is associated with their strategy use on the Ramp question.

Coding Responses on the Written Posttest

The Ramp item first asked the students to pick which balls to compare in order to figure out what makes a ball go fast down the ramp. There was no specification of how many balls to compare, although most (79%) choose to compare two. We coded the responses to the Ramp Comparisons task as CVS, GPS, or Neither. A student using the Control of Variables Strategy should pick two cases to compare that vary only on one characteristic (weight, shape, or size). There are two possible pairs for which this is the case: tennis ball & baseball or soccer ball & basketball. If the student chose either of these pairs, their response was coded "CVS".

If students are using GPS, they should pick cases that have the same outcome (time down the ramp) then look for what characteristics are common across these cases. In the case of the Ramp Comparisons task there are two ways to pick a common outcome: (i) pick the fastest balls (baseball & bowling ball), which took 1.5 seconds, or (ii) pick the slower balls (tennis ball, soccer ball, & basketball), which took 2 seconds. If a student responded with either of these groupings, their response was coded as GPS (1).

The second question of the Ramp item, which we will refer to as the Ramp Conclusions task, asked students to use the data to decide what affects the time needed for the ball to roll down the ramp. There was no specification of how many factors could be affecting the speed, but most (80%) put only one. Using either strategy should lead to the same (counterintuitive) conclusion for this data, which is "shape" (2). The students' responses were coded as correct if they identified shape (hollow or solid) and did not list any other characteristics.

Two coders independently coded 20% of the responses, which were randomly selected from the posttests. They agreed on 100% of the codes for both Ramp questions before discussion.

Results

Inquiry Strategy on The Ramp Question

Did the students in each condition learn to apply the strategy to the new physics topic? Figure 3 shows a histogram of the strategies used on the Ramp Comparisons task, by condition. Note that the strategy used on this question tends to align with the condition. A Chi-square test of independence shows that this association is significant $\chi^2(2, N = 100) = 6.14$, p < .01. Also note that in the GPS condition (N=53) there were a relatively large proportion of students who used CVS. The converse is not true: in the CVS condition (N=47) only one student used the GPS strategy. This suggests that many students in GP (and by implication, CV) may have already been familiar with CVS. Lastly note that in both conditions (but especially in the GP condition) there is a fairly large proportion of "Neither" codes, i.e., comparisons that did not conform to either strategy.

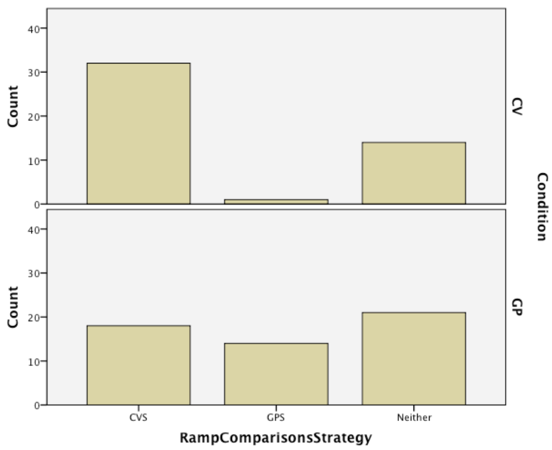


Figure 1. Histogram of inquiry strategies used on the Ramp Comparison task, by condition.

Were the students able to draw the right conclusions from their comparisons? Table 1 is a contingency table for the strategy used on the Ramp Comparisons task with correct responses on the Drawing Conclusions task. There is a significant association $\chi^2(2, N = 100) = 7.33, p < .05$ between using either strategy on the Making Comparisons task and drawing the correct inference in the Drawing Conclusions task. However, Chi-square does not isolate which interaction is driving the effect. To test whether the large proportion of "Neither" codes in GP was behind the association, we collapsed CVS and GPS into a single category "Either" and found that the association was still significant $\chi^2(1, N = 100) = 7.29, p < .01$. This suggests that students using either strategy were more likely to draw the correct conclusion from their comparison.

The association of drawing the correction conclusion with condition was not significant $\chi^2(1, N = 100) = .17, p > .05$. This could be due to the low incidence rate of people using GPS (n=15), the high number of students in the GP condition using CVS, and the high incidence rate of "Neither" codes. All together, the results suggest that both strategies were helpful for those that used them, but that in future iterations instruction should focus on improving the uptake of GPS.

Table 1: Contingency table for Ramp Comparisons strategy and correct/incorrect Ramp Conclusions.

Ramp Comparisons strategy:	Ramp Conclusion: incorrect	Ramp Conclusion: correct	Total
CVS	12	38	50
GPS	4	11	15
Neither	18	17	35
Total	34	66	100

Strategy Choice and Performance on the Balance Act

To complete the Balance Act Challenge mode, students needed to successfully predict eight challenges in a row. The students were coded for completing the Challenge mode or not. Table 2 is a contingency table for the strategy used on the Ramp Comparisons task with their successful completion of the Balance Act Challenge mode. The association between strategy used on the Ramp Comparisons ramp question and completion of the Balance Act Challenge mode is significant $\chi^2(2, N = 100) = 6.14, p < .05$. The table shows that of the students who used GPS, more than half (60%) completed the Balance Act Challenge. Compare this with students who used the CVS strategy, of which slightly less than half (46%) completed the Balance Act, and with the students who used neither strategy, of which only 26% completed the Balance Act. The table shows that students who used either strategy on the ramp task had a better chance of completing the challenge than those who did not use either strategy.

Table 2: Contingency table for Ramp Comparisons strategy and Balance Act completion.

Ramp Comparisons Strategy	Balance Act challenge not completed	Balance Act challenge completed	Total
CVS	27	23	50
GPS	6	9	15
Neither	26	9	35
Total	59	41	100

We also coded students' performance on the Balance Act according to the maximum number of challenges in a row they got correct. Those who completed the challenge have a max score of 8, and the overall mean was 6.59 (see Figure 4). We find that the strategy used on the Ramp Comparisons task is also associated with the maximum number of challenges won in a row in the Balance Act Challenge mode. An analysis of variance with the maximum score on the balance scale crossed with the inquiry strategy used on the ramp question was significant $F(2,98) = 6.143, p = .003$. The means of the maximum balance scale challenge for those who used either GPS or CVS on the ramp question were significantly higher than those who used neither strategy. The means for GPS are descriptively higher than for CVS, but this difference does not rise to significance $t(63) = 1.08, p = .29$.

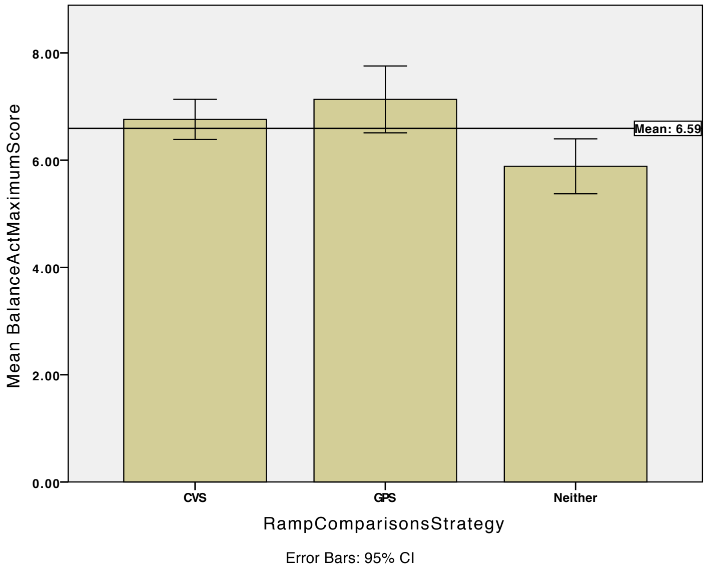


Figure 2. Bar chart comparing means of the maximum number of challenges answered correctly in a row on the Balance Act, by strategy used on the ramp question.

Conclusion & Discussion

The written posttest item assessed whether students learned to apply and draw correct inferences from the instructed inquiry strategy (CVS or GPS) in a new physics context. Analysis shows that the strategies used on this question were associated with each respective condition (CV or GP), and that using either inquiry strategy was significantly associated with finding the correct answer. The computer-based posttest assessed whether students figured out a multiple variable relationship of torque in the context of balancing. Analysis suggests that if they used either strategy on the written posttest, they did significantly better at picking up the multivariate relationship (weight x distance) on the computer-based posttest. This is suggestive that the strategies do help them figure out multiple variable relationships.

In this preliminary study, we found evidence that GPS is at least as effective as CVS for figuring out how multiple variables interact with each other. There is also a hint that GPS helps more than CVS, although due to the low incidence of GPS strategy use, the mean difference did not rise to the level of statistical significance. Overall, this study serves as an existence proof that it is possible to explicitly teach the strategy of seeking a general principle to middle school students in ways that help them figure out multiple variable interactions on their own. This is an important implication for instruction.

Further research is needed to verify that students really are using these strategies as they explore with the Balance Act simulation. This could be corroborated by examining their work on Balance Act either through the session log files and/or by videotaping their screen as they work with the PhET, coding when their moves are consistent with GPS or CVS. The lack of significant associations by condition with Ramp Conclusions and Balance Act measures, due to the low overall incidence rate of students using GPS, as well as the high rates of students using CVS or neither strategy, suggests that future studies should focus on exploring ways of teaching GPS more effectively.

Lastly, future work will focus on finding the productive common ground between students' inquiry into related variation and inquiry into the meaning of concepts. diSessa (2008) has described these modes of inquiry as distinct but complementary, pointing out that few studies have examined their intersection. Ideally, exploring related variation in realistic contexts could inform students' understanding of the meaning of the related contexts, for example, by making sense of the counterintuitive conclusions drawn from their application of inquiry strategies.

Endnotes

(1) There is an ambiguous case: if students chose the soccer ball and the basketball, they could either be controlling variables by making a comparison based on the one characteristic being different (weight), or seeking a general explanation based on the common outcome (2 seconds). We therefore took a conservative approach to coding for GPS; for the slow (2 sec) balls, *all three* had to be compared in order to be coded as a GPS response. If just the basketball/soccer ball pair was selected, it was coded as CVS. It is conservative in that it may be throwing out GPS responses, thereby weakening the correlation between the instructional condition and the strategy used on the ramp

task. Anyone comparing this particular pair is much more likely to be doing so based on the CVS strategy, since that is one of 2 possible correct CVS pairs, while for the GPS strategy that would be one of 4 possible GPS pairs. There were only 5 cases of this, and 3 additional GPS pairs that were excluded based on the conservative coding scheme..

(2) Even though most people intuitively expect the mass and the size of the shapes to matter, it turns out that mass and radius both cancel out of Newton's equations (in the limit of low rolling speeds). Hollow objects have greater rotational inertia, and so cannot speed up as quickly, no matter what the mass or the size is.

References

Chase, C., Shemwell, J., & Schwartz, D. (2010, June). Explaining across contrasting cases for deep understanding in science: An example using interactive simulations. *Proceedings of the 9th International Conference of the Learning Sciences-Volume 1* (pp. 153–160). International Society of the Learning Sciences.

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., Dohmen, I., Shemwell, J., Chin, D., & Chase, C. (2012, April). Seeing the forest from the trees: A comparison of two instructional models using contrasting cases. *Paper presented at the 2012 American Educational Research Association Annual Conference* (pp. 1–10). Retrieved 3/5/2014, from the AERA Online Paper Repository.

diSessa, A. (2008). A "theory bite" on the meaning of scientific inquiry: A companion to Kuhn and Pease. *Cognition and Instruction 26*(4), 560-566. doi:10.1080/07370000802391760

Ford, M. J. (2005). The game, the pieces, and the players: Generative resources from two instructional portrayals of experimentation. *The Journal of the Learning Sciences*, *14*(4), 449–487.

Kuhn, D. (2007). Reasoning about multiple variables: Control of variables is not the only challenge. *Science Education*, *91*(5), 710–726. doi:10.1002/sce

Kuhn, D., Pease, M., & Wirkala, C. (2009). Coordinating the effects of multiple variables: A skill fundamental to scientific thinking. *Journal of Experimental Child Psychology*, *103*, 268–284. doi:10.1016/j.jecp.2009.01.009

Shemwell, J., Chase, C., & Schwartz, D. (under review). Seeking the general explanation: Revisiting the wisdom of Francis Bacon.

Wieman, C., Adams, W., & Perkins, K. (2008). PhET: Simulations that enhance learning. *Science, 322*(5902), 682-683. doi:10.1126/science.1161948

Acknowledgments

This work was funded by NSF 09-602, and The Hewlett Foundation. The authors would also like to thank Doris Chin, Kristen Blair, Joe Prempeh, Min Chi, and Michael Ford for valuable assistance with the study. Additional thanks go to Kathy Perkins, Michael Dubson, and the PhET team.