A Comparison of Two Methods of Active Learning in Physics: Inventing a General Solution versus Compare and Contrast

Doris B. Chin, Min Chi, & Daniel L. Schwartz

Stanford Graduate School of Education

Instr Sci Accepted 22 April 2016 DOI 10.1007/s11251-016-9374-0

ABSTRACT

A common approach for introducing students to a new science concept is to present them with multiple cases of the phenomenon and ask them to explore. The expectation is that students will naturally take advantage of the multiple cases to support their learning and seek an underlying principle for the phenomenon. However, the success of such tasks depends not only on the structure of the cases, but also the task that students receive for working with the examples.

Two studies used contrasting cases in the context of teaching middle-school students about projectile motion. Using a simulation and the same set of cases for all students, students completed a traditional "compare and contrast" approach, or an instructional method called "inventing," where students try to produce a single general explanation. The results show that inventing led to superior learning. Examination of student worksheets revealed that the "compare and contrast" instruction led students to focus mostly on the level of discrete, surface features of the phenomenon. Rather than trying to account for the variation across cases, students simply noticed each instance of it. In contrast, the inventing task led students to consider how the variations across the cases were related. As a result, "invent" students were more likely to search for and find the unifying functional relation. Driving towards an overall explanation is a fundamental tenet of science, and therefore, it is worthwhile to teach students to do the same.

Key words: Science education, science instruction, inventing, compare and contrast, contrasting cases

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Education IES grant R305A140314. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the granting agencies.

INTRODUCTION

Accumulating evidence indicates that “active learning” (Bonwell & Eison, 1991), such as the use of small group problem-solving tasks in classrooms, produces better learning outcomes than simply listening to lectures alone (Prince, 2004; Freeman et al., 2014). Moreover, by engaging students more deeply, active learning techniques can prepare students to learn more from *subsequent* lectures by rendering concepts more immediate or relevant (Schwartz & Bransford, 1998). The growing body of evidence and increasing use of active learning naturally leads to the question of how to design good experiences for active learning (for a thorough, differentiated taxonomy of learning activities, see Chi, 2009).

A ubiquitous form of active learning found across many disciplines is to give students sample cases and ask them to compare and contrast the different examples. Previous research indicates that asking people to find similarities and differences across multiple, juxtaposed cases or problem instances can positively influence learning and transfer, including work on analogy (Gentner et al., 2003), perceptual learning (Biederman et al., 1987; Marton & Booth, 1997), memory (Bransford, Franks, Vye, & Sherwood, 1989), procedural learning (Rittle-Johnson et al., 2009), and category induction (Williams et al., 2010). Indeed, a meta-analysis by Alfieri, Nokes-Malach, and Schunn (2013) computed that case comparison yields an average gain of 0.5 sigma compared to lessons that do not include case comparison.

As an example of a simple compare and contrast activity in physics, consider students first being introduced to the concept of buoyancy. These students might be asked to explore various objects floating and sinking in different liquids. In a very simplified control of variable strategy, students could explore the question of what makes things float, by identifying similarities and differences across the different examples, such as size, mass, shape, material, and liquid.

However, simply cataloging the easily identifiable features is insufficient for understanding buoyancy. This is true for most physics concepts, because the important similarities and differences involve “invisible” functional relations among multiple features. Functional relations refer to how variables interact to produce a result, as in force = mass X acceleration and density = mass ÷ volume. For instance, a plank of wood and an ingot of steel look and feel different, and one floats on water while the other sinks. If students confined their analysis to the perceptible properties (surface features), they would not find the functional relation (deep structure) of relative densities, which is what determines whether one object floats and another sinks in a particular liquid.

In this paper, we present an approach to designing comparison-based, active learning tasks that help students discern functional relations across multiple cases. First we briefly describe the underlying theories from cognitive and perceptual psychology that guide the selection of cases. Second, we consider the tasks students might receive when given the cases. Finally, we describe two studies that isolate the effect of the task orientation that students receive. They show that the typical directive to “compare and contrast” produces an *instructional backfire*. It leads students to notice what they already know, even though the point of doing compare and contrast activities is to help students discover what is new to them. In the alternative, if students are asked to invent a single explanation for the similarities and differences across the cases, they discover the underlying functional relation, at least when provided with well-designed contrasting cases.

The Selection of Cases for Compare and Contrast Activities

Case comparison is a form of inductive activity where people find features and patterns that they have not recognized before. There are two variants of inductive learning that offer different proposals for how to select cases. The first comes from the cognitive literature on category formation and analogical learning. It emphasizes finding a general pattern or rule from a set of instances. Fried and Holyoak (1984), for example, demonstrated that both the central tendency and the spread of a distribution influenced how experimental subjects categorized different 10x10 grid patterns. Gick and Holyoak (1983) explored whether similar or disparate story analogs improved schema induction. Rittle-Johnson and Star (2007, 2009) found that students developed more procedural flexibility in algebraic manipulations if they compared alternative solutions side-by-side versus sequentially. General conclusions about the nature of the stimuli that make induction possible are that two examples are better than one, and the examples should be disparate and side-by-side, so that students can best isolate the common deep structure while recognizing the irrelevance of the surface features (e.g., Christie & Gentner, 2010).

The second variant of inductive learning comes from perceptual psychology (e.g., Gibson & Gibson, 1955). Biederman and Shiffrar (1987), for example, created a set of cases that brought novices in sexing day-old chicks to near expert levels within a matter of minutes. The cases helped the employees see the features that experts see; they improved discernment (e.g., Bransford, et al., 1989; Eisner, 1972). Here, the prescription for the design of cases is to make them as similar as possible, so learners can identify the distinctive features. For instance, Gick and Paterson (1992) demonstrated that the inclusion of "near misses" – contrasts that differ on a single dimension – improve schema induction and transfer.

[Insert Figure 1 about here]

Figure 1 summarizes the prescriptions of the two traditions for helping people to learn. In one case, it is important to pick disparate cases so people can sort out the deep structure from the surface features, whereas in the other, it is important to pick near-miss cases so people can isolate critical features. To resolve this seeming conflict, it is important to design cases that provide both similarities and differences. Additionally, to help students find the functional quantitative relations, the cases need to provide systematic variation across the relevant variables. Thus, there are two principles. First, subsets of cases need to share common values, while differing on others, to provide students with controlled variation. Second, the cases need to exhibit parametric variation on the key properties that provide data for students to learn how more and less of each property matters for the behavior of the other properties. We provide a concrete example.

[Insert Figure 2 about here]

Figure 2A shows the set of contrasting cases used for the current studies with middle school children learning projectile motion. The cases exhibit systematic variation for inducing how time aloft and speed determine the range of a horizontal (0° angle) shot. First, the cases vary initial height (hang time) and speed, and therefore, the range of the shot. Second, they hold height or speed constant across several cases, while the range of the shot varies. Ideally, this helps students notice that height and speed jointly determine the range. Note that the cases do not include any variation in the angle of the launch, so these cases would be inadequate for learning about the effect of angle. (They are for adolescents who do not know curvilinear functions, so height, and thus time aloft, stands in as a simpler proxy for the effects of gravity on projectile motion.)

Task Orientation

The cases in Figure 2 have been designed to improve student chances of inducing the key functional relation for projectile motion. They are similar in style to cases used to support induction of the ratio relation in density (Schwartz, Chase, Oppezzo, and Chin, 2011) and concepts of variance in statistics (Schwartz & Martin, 2004). However, despite the care given to the cases, we propose the instructional task students receive is also critically important for learning. In traditional compare and contrast activities, the simultaneous presentation of cases helps alleviate the common problem that students treat sequentially presented examples as unrelated and fail to notice their defining features (e.g., Rittle-Johnson & Star, 2009). However, simple compare and contrast activities may introduce a different problem, especially for novices. We propose that a major risk of a compare and contrast task is that it plays into the tendency of novices to rely on discrete and easily perceived properties (Gibson & Gibson, 1955). For example, Chi, Feltovich, and Glaser (1981) demonstrated that undergraduates, who had completed a semester of mechanics, categorized mechanics problems based on perceptible surface features. Asking total novices to list similarities and differences should yield the same problem, even more so. For example, students may notice that the projectile cases have same and different heights, and same and different velocities, and be satisfied that they have completed the assigned task. So, although students are engaged in active learning with compare and contrast, they are engaged in the wrong activity for learning functional relations and generalizable concepts. This has large implications for transfer and students' abilities to solve new, related problems in the future.

Our alternative to traditional compare and contrast activities is to ask students to *invent* a single, general solution that handles all the cases (Schwartz, et al., 2011). With the projectile cases above, students receive the fanciful cover story of an amusement park that shoots people from cannons (Figure 2B). Their task is to use the cases as data and to invent a way to place the landing pool so that visitors get a good landing, *no matter what combination of height and speed* *a visitor chooses*. We predicted this instructional task would drive students to search for an underlying functional relation that combines quantities, because no discrete quantity can solve the problem. More generally, encouraging students to search for a general solution is critical. Whereas general solutions and explanations are a cornerstone of physics, people do not naturally seek them. For example, researchers have found that without the directive to produce a general explanation, college students explained multiple examples of inducing a current with a magnet as independent problems rather than taking them as instances of the same underlying principle (Shemwell, Chase, & Schwartz, 2015). In the context of projectile motion, Maloney (1988) warned that student "rule usage was quite flexible with essentially no consideration of the fact that all of the situations involved the same type of motion. That is, the subjects seemed to treat each situation as unique with no need to correlate a rule on one task set with the rules on related tasks sets" (p. 511). The hypothesis driving the current work is that the instruction to invent a general solution can overcome the natural tendency to treat features and problems discretely, which in turn, can help students find the target functional relations in a set of contrasting cases.

RESEARCH OVERVIEW

To test our assumptions on the importance of task orientation, we conducted two studies with adolescents to compare the effects of two task instructions: (a) compare and contrast the cases versus (b) invent a general solution. Both treatment groups received *identical* contrasting cases. We examined students' learning outcomes, as well as their worksheet answers to clarify the problem solving that led to the different learning outcomes.

The lessons were organized around the PhET "Projectile Motion" simulation. Learners can fire objects out of a cannon by manipulating different variables, including angle, height, speed, mass, and air resistance (Figure 3). To simplify the domain for middle school students, we focused on the horizontal x-component of shots and removed air resistance. This way the horizontal distance (d_x) is given by the horizontal speed (s_x) multiplied by time aloft: $d_x=s_x*t$.

[Insert Figure 3 about here]

Figure 4 shows the general procedure of both experiments, Study 2 being a replication of Study 1. Both studies included three phases. In Phase 1, students familiarized themselves with projectile motion and the simulation. Phase 2 contained the sole experimental manipulation, and students worked with identical 0° angle cases shown in Figure 2 rather than the simulation. They were all given the task to “put the water tank in the right place for each visitor,” but half used the cases with a Compare and Contrast orientation (CC), and the other students used the cases with an Invent orientation (Invent). Finally, in Phase 3, students again used the simulation, this time during a predict-observe-explain (POE) activity.

[Insert Figure 4 about here]

The studies tested three specific hypotheses. The first involved the Phase 2 worksheets. We hypothesized that students in the CC condition would tend to list discrete similarities and differences and fail to notice relations among features. In contrast, students in the Invent condition would tend to identify relations among features and not bother listing obvious surface comparisons.

The second hypothesis involved the mid-test, immediately after the Phase 2 treatment.

Students received a word problem that gave the parameters for a new cannon shot, and they had to determine the landing point. We expected the Invent students to be better at predicting the landing point for this new problem, because they had paid attention to the functional relations in Phase 2. In contrast, the CC students would not have paid attention to the functional relations during Phase 2, and therefore they would be worse at generalizing to a new problem.

Finally, we hypothesized that the activity of inventing a general solution would better prepare students for future learning (Schwartz & Bransford, 1998). Previous studies have demonstrated that inventing with contrasting cases helped students notice key relations among quantities, and thus, even if they never invented a full-blown, correct solution on their own, they were better prepared to understand an exposition of that solution (Schwartz & Martin, 2004). In the current experiments, we wanted to determine whether inventing would prepare students to learn from a very commonly used learning tool in physics, an interactive simulation, as opposed to an expository treatment. The learning opportunity came during Phase 3, in the form of using the simulation to predict the landing point of four projectile cases using a Predict-Observe-Explain (POE) framework. To determine if students learned from the POE activity, we looked for improvements from the mid-test to an isomorphic post-test. Our basic assumption was that students who noticed relations among quantities during Phase 2 would be more prepared to learn the multivariate, functional relation underlying the simulation in Phase 3. Because we hypothesized that the Invent task would lead students to notice the multivariate relations more frequently than the CC task, our third hypothesis was that Invent student would learn more than their CC peers from the POE activity.

STUDY 1

Methods

Participants

Two 6th-grade advanced math classes with the same teacher participated. The two classes were from different schools in the same high SES school district (Caucasian 51%, Asian 29%, Hispanic 8%, Other 12%: Free-Reduced Lunch Program 4%; ELL 9%). Logistics required that the intact classes be randomly assigned to one of two treatments, Compare and Contrast (CC, n = 21) or Invent (n = 16). There were no significant differences between the two groups on any teacher measures just prior to the study: Class test scores ($F_{(1,35)} = 0.21$, MSE = 39.34, $p = 0.65$) and homework and quiz scores ($F_{(1,35)} = 0.12$, MSE = 20.25, $p = 0.74$).

Design and Procedures

Instruction occurred on three consecutive Friday classes (50 min each). In Phase 1, students worked individually to familiarize themselves with the simulation. They explored different ways to hit targets and had to determine the meaning of the black plus signs **(+)** on the shot path (i.e. the object's position at each second).

In Phase 2, students worked in pairs, per regular classroom practice. Both treatment conditions received the same cover story (Figure 2B), and exactly the same set of contrasting cases (Figure 2A). All groups were told to figure out the right place for the water tank such that each visitor has a good landing. The treatment difference occurred only in the subsidiary instructions students received. In the service of finding the right place to put the water tank for each visitor, CC pairs were told to "*Compare and contrast the examples and companies. Explain the similarities and differences*." Invent pairs were told to "*Invent a single method to figure out* *where to put the pool no matter which company and speed a visitor chooses*." The worksheet activity lasted 15-20 minutes. Afterward, students individually answered a brief mid-test item.

In Phase 3, students again worked individually with the PhET simulation, this time on a POE activity. Students predicted the range of the four cases shown in Figure 5 and checked their predictions with the simulation. They also explained how they made their predictions and how their prediction could be improved, if wrong. Students then took a brief post-test.

[Insert Figure 5 about here]

Measures and Coding

The mid-test and post-test were each a single isomorphic word problem. An answer was correct only when the student provided both an accurate range and an appropriate explanation. Here are the two word problems:

Mid-test Word Problem:

A bowling ball is shot out of a cannon at a speed of 18 m/s. The cannon is at a height of 11 m. The ball has a mass = 9 kg and diameter = 0.3m. The ball lands in 1.5 seconds. There is no air resistance.

a) How far does the ball go before landing (what is its range)? _______________

b) How do you know?

Post-test Word Problem

A cannon is on top of a building 78 m high. It is aimed straight out across a flat field of grass. A large pumpkin is shot out of the cannon at a speed of 18 m/s. The pumpkin has a mass = 20kg and diameter = 0.3m. It is in the air for 4 seconds before it lands. Assume there is no air resistance.

a) What is the range of the pumpkin (how far does it go before landing)? _____________

b) How do you know?

We also coded the student worksheets from Phase 2. We were interested in whether students described similarities and differences in terms of zero, one, two, or three physical dimensions simultaneously. Depending on the degree to which students found functional relations among the three variables (height, speed, and distance), we coded student statements as single-, double-, or triple-factor answers. Table 1 shows examples of student statements and their codes.

[Insert Table 1 about here]

We counted the number of unique single-, double-, and triple- factor statements in each group's worksheet answers. For example, Figure 6 shows one CC pair's statements. Among them, both "same m/s (10 m/s)" and "same m/s (8 m/s)" indicate speed, and therefore, the students only received credit for one single-factor (i.e., speed) for these two statements. One researcher coded all the worksheets. A second researcher coded a random subset of 40% of the worksheets. Agreement was above 96%.

[Insert Figure 6 about here]

Results and Analysis: Study 1

The results were definitive. As predicted, the Invent students found the underlying functional relation, whereas the CC students noticed discrete features. The first piece of evidence is the types of factors they listed on the worksheet in Phase 2. A triple-factor means that a team identified a functional relation of three quantified dimensions (e.g., time x speed = range), and a single-factor means that a team identified a single quantitative dimension (e.g., speed is the same across two cases). Table 2 holds the percentage of groups who noted each type of factor. Every Invent group stated a triple-factor, and every CC group identified single-factors. Moreover, the CC groups averaged 3.0 separate single-factors (SD = 1.1). Thus, the CC groups successfully noticed three variables, but they did not put them into a functional relation, whereas the Invent groups did.

[Insert Table 2 about here]

One possibility is that the CC students noticed triple-factors, but they did not include them in their tabulation of similarities and differences, given that there was no explicit requirement to do so. The mid-test suggests this was not the case for the majority of CC students. Solving the mid-test item required an understanding of the key triple-factor. Figure 7A indicates the Invent students greatly outperformed their CC peers: 87.5% versus 38% correct respectively; $p = .003$, Fisher's Exact Test.

[Insert Figure 7 about here]

A final question is whether either condition prepared students to learn from the POE activity in Phase 3. The Invent students were already close to ceiling on the word problem, whereas the CC students had room for improvement. The post-test performance after the POE activity (Fig. 7B) indicates that the CC students did not learn from the POE activity, whereas one further Invent student learned the functional relation determining the range. Thus, the CC activity had not prepared the students to learn from further inquiry, whereas ceiling effects prior to the task preclude any conclusions for the Invent students.

Discussion

The study tested whether CC activities lead students to notice discrete features whereas Invent activities lead students towards underlying functional relations. The key evidence comes from an analysis of student worksheets during the activity of analyzing the contrasting cases. The results showed that 100% of the Invent pairs made triple-factor statements that captured the functional relations among the variables, whereas 100% of the CC groups identified obvious discrete features. These differences in what students noticed had large implications for their subsequent individual problem-solving performance. The Invent students were more accurate than their CC peers on the quantitative word problems at both mid-test and post-test. Finally, the CC activity had not prepared students to learn from the simulation during the subsequent POE phase, whereas the Invent students were already near ceiling, so it is unclear what they may have gained from the simulation.

A major methodological limitation of the current study was the assignment of intact classes to each treatment. While the CC and Invent students had comparable levels of incoming achievement, it is possible that some unidentified advantage occurred in one class but not the other. Therefore, we replicated the study by randomly assigning students to treatment within their classes instead of between classes. We also addressed two other questions with Study 2.

We assume that the Invent activity works by helping students find the functional relation, and finding the functional relation is what improves learning and subsequent test performance. However, we did not demonstrate that causal relation. An alternative, though unlikely hypothesis, might be that it is not the discovery of the functional relation *per se* which serves as the active learning ingredient, but rather the more engaging nature of inventing. Engaged students try harder and learn better (Fredericks, Blumenfeld, & Paris, 2004; Finn & Zimmer, 2012). To demonstrate the causal role of finding the functional relation, we need to show that CC students who find the functional relations during the worksheet activity also do well on the tests. However, because only one pair in the CC condition found the functional relation during the worksheet phase, we could not make this demonstration. Therefore, in Study 2, we made a small amendment to the CC instructions to try to increase the number of students who found the functional relation, so we could then partial the contributions of the inventing task versus discovering the functional relations on test performance. The amendment was to include a sentence that told students to notice that each company always uses the same height of launch, which we thought might lead the CC students to consider differences within a context of similarities.

A final consideration is the lack of any effect of the POE activity on the CC students' learning. One possible interpretation is that the simulation is too complex or confusing for middle school students, so they could not learn from it, no matter what kind of instruction came earlier. To address this issue, we modified the Phase 3 instruction to include two, *separate* rounds of POE, a pair of cases at a time, in the hope that the iteration would promote deeper reflection and greater learning. In addition, we added an assessment item; students completed a projectile motion problem from the Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992) before and after the first time they used the simulation, but before any treatment differences. We used the item that asks students to choose the correct trajectory of a 0° angle projectile, which is notoriously problematic for even college students. If students improve on this item, it would show that students can learn from the simulation, and the lack of POE effect was not due to a problem with the simulation.

STUDY 2

Study 2 was a near replication of Study 1 with new students one year later. We made three primary modifications: random assignment to treatment within class; small adjustments to the CC and POE worksheets to increase CC performance; and, the administration of an FCI task that required students to choose the trajectory of a projectile, before and immediately after Phase 1. The Phase 3 POE session was again uninformative, as only a single student (CC condition) improved on the word problem test. For the sake of brevity, we leave out the Phase 3 data.

Methods

Participants

The study involved similar students from the same school district a year later. Students were assigned within class to the two treatments, Invent (*n* = 17) or Compare and Contrast (CC, *n* = 17). Stratified random assignment successfully balanced the Invent and CC conditions; there were no significant differences between conditions across five types of teacher measures, including students' most recent test scores prior to the study ($F_{(1,32)} = 0.06$, MSE = 22.90, $p = .80$).

Materials and Procedures

Study 2 covered the same lessons on projectile motion as Study 1, but condensed the study to two Friday sessions (50 min each), two weeks apart, in the winter trimester. As before, students completed worksheets in groups (8 groups per condition), but worked independently at all other points. Here, we only describe the variations from Study 1.

Two months prior to Phase 1, students completed the projectile motion, Force Concept Inventory (FCI) item #16 (Hestenes, Wells, & Swackhamer, 1992). The temporal separation of the pre-test from the study was to reduce test reactivity (i.e., we did not want students coming into the simulation looking for the answer to the test item). There were no intervening lessons on projectile motion. Then, after Phase 1, in which students worked with the simulation and learned the meaning of the + signs on the projectile path, students completed the FCI item again.

We made a small change to the CC worksheet by including the line, “Each company always uses the same height.” The hope was that this would lead more CC students to notice a difference within a context of paired similarity, rather than just comparing across all six cases and haphazardly picking similarities and differences.

Finally, because of the in-class random assignment, all students were led through the cover story of finding the right place to put the pool for visitors, but then were simply told that their task was listed on second worksheet page and to read the instructions carefully. For Study 1, the intact classes allowed the instructor to read all of the instructions together with the students.

Results

The simple summary is that inventing again led more students to find the underlying functional relation than the compare and contrast (CC) condition, and finding the functional relation, regardless of condition, led to the superior test performance.

In Study 2, all of the Invent groups generated triple-factor statements during Phase 2, whereas half of the CC groups did. Thus, we succeeded at improving the CC performance, and the Invent students remained at ceiling. The remaining four groups (8 CC students) only generated single-factor statements.

[Insert Figure 8 about here]

Figure 8 provides the relevant results for separating out the effects of treatment and triple-factor discovery on test performance. Inventing led to more triple-factor generation than CC (100% vs. 53%, respectively), and triple-factor generation led to better test performance than non-triple-factor generation (85% vs. 25%, respectively). To determine statistically whether the association between triple-factor generation and test performance differed by condition, a loglinear model was fit to the crossed factors of Triple-Factor Generation (Triple, Non-triple), Test Performance (correct, incorrect), and Condition (Invent, CC). Backwards elimination found that the 3-way interaction effect was not significant, χ^2 (1, N = 34) = 0.001, $p = .98$, indicating that the benefit of generating a triple-factor statement for test performance did <u>not</u> vary by condition. At the next step, the 2-way interaction effect of Condition X Test was found to be non-significant, χ^2 (1, N = 34) = 0.4741, $p = .49$. The final step found that the remaining two-way interactions were significant for the fit of the model. Invent students were more likely to generate a triple factor statement, χ^2 (1, N = 34) = 13.59, $p < .001$; and, successful triple-factor generation was significantly associated with successful performance on mid-test quantitative item, χ^2 (1, N = 34) = 9.87, $p = .002$. Thus, the model that included the higher-order interactions of Condition X Triple-Factor Generation and Triple-Factor Generation X Test Performance, gave the best fit to the data (Likelihood Ratio χ^2= .475, df = 2, $p = .789$).

In total, these analyses indicate that, regardless of condition, whether or not a student generated a triple-factor statement during the worksheet phase was a significant predictor for students' quantitative problem solving performance and they indicate that the Invent group outperformed their CC peers *because* the former generated more triple factors than the latter.

It is worth noting for instructional purposes that in both studies, although 100% of the Invent *teams* got the triple-factor on the worksheet in Phase 2, not every Invent student answered the subsequent midtest question correctly (two students for both study 1 and study 2 answered incorrectly). Conversely, a small number of the CC students who did not generate the triple factor statement with their partners on the worksheet, *did* answer correctly on the midtest (six students for Study 1 and two for study 2 respectively). There are several possible explanations for such phenomena: for example, the Invent students who failed to correctly answer the midterm question were those did not come up with the triple-factor themselves, but rather their partners did during the worksheet phase; for the CC groups, some CC students may have actually noted the triple-factor relationship, but did not write it down on their group's worksheet. It is difficult to determine the exact cause(s) for these anomalous students, however further research could utilize student interviews or videotape to shed light on students' cognitive processes during contrasting cases work and provide insightful information for educators to improve implementation of such activities.

A subsidiary question addressed by the study was whether students could learn from the simulation, given the lack of POE effect in Phase 3 of Study 1. At pre-test, 61.1% of students answered the FCI item correctly. At the end of Phase 1, immediately after using the projectile motion simulation for the first time, 77.8% of students were correct, a 43% gain in possible improvement (Fisher's Exact Test, p =.036). This improvement is comparable to those exhibited by the high school and university students in the original populations studied by Hestenes, Wells, and Swackhamer (1992), which ranged from 38-71% of possible gain after a semester or more of physics instruction. Thus, it appears that students can learn from the simulation, even with minimal instruction. However, it remains an open question of how to help them learn underlying functional relations from the simulation, given that POE was ineffective in both studies. An approach that structures an inventing task for the simulation may prove effective.

CONCLUSION

In science classrooms using active learning, students often complete tasks where they explore data, and then draw conclusions from those data. When there are data from more than one instance, the expectation is that students will take advantage of the multiple instances to support their learning. The success of these multi-case tasks will depend on both the structure of the cases, and the task that students receive for working with the cases.

In the studies presented here, we emphasized the task students receive. We compared two instructional approaches to using contrasting cases in the context of adolescents learning projectile motion: a canonical “compare and contrast” approach and a technique we call “inventing.” Given identical contrasting cases, the results show that the inventing instructional treatment led to superior learning outcomes over a compare and contrast condition. Examination of the students’ contrasting cases worksheets revealed that the directive to “compare and contrast” led students to focus mostly on the level of discrete, perceptible features. Rather than trying to account for the variation in the cases, they simply noticed each instance of it. The inventing task, which asked students to find a single general explanation, led students to consider how the variation across the cases is related. As a result, these otherwise identical students were more likely to search for and find the functional relation that made the cases variations of the same underlying principle. Previous studies conducted under the framework of expertise development have found that, in many disciplines, novices tend to represent disciplinary knowledge as easily identifiable discrete surface features, while experts are characterized by representing domain knowledge as connected, underlying functional relations among multiple features (Alfieri, Nokes-Malach, & Schunn, 2013; Bransford, Brown , & Cocking , 1999; Hatano

& Inagaki, 1986). The results here indicate that students who were under compare and contrast condition seemingly demonstrated more novice-like behaviors, while students under the inventing condition developed more expert-like characteristics. Thus, asking students explicitly to generate a general function can be an important instructional strategy for facilitating expert-like learning behaviors which will result in greater learning outcomes.

The invention instructions could be interpreted as asking students to find an analogy across cases: they are supposed to figure out the deep structure that makes all the cases an instance of the same principle. However, glossing the inventing task as analogical reasoning dismisses some of its unique properties. Though both analogy generation and invention do involve relational thinking, analogical reasoning is insufficient to understand a functional relation. Analogical reasoning typically involves inducing qualitative relational similarities across diverse examples, such as current is like a teaming crowd of moving people. The surface features of the instances disappear in the abstract formulation of the analogy (e.g., the analogy would work if it were to a teaming school of fish). For functional relations, such as speed, the variation in the surface features is precisely what the functional relation seeks to explain, and quantitative precision is necessary for students to develop an account that explains the variation in specific surface quantities, such as height, time, and range. On the one hand, this would seem to be an even more demanding task than finding an analogy. On the other hand, people can recruit mathematics to help find and describe the relational structure, which is rarely the case when formulating an analogy.

There is a delicate balance in helping students find functional relations. Students need to delve beyond surface features to find the underlying relational structure, however they must also attend to the surface features themselves, as these factors hold the relevant quantitative information. Providing students with contrasting cases is one way to help students strike that balance, but, as we showed here, it also runs the risk of leading students to treat surface features as discrete entities. Thus, while the cognitive task of inventing differs in important ways from analogical reasoning, they both confront the same problem. Students may not spontaneously search for a common deep structure across instances without prompting (for a review, see Richland & Simms, 2015).

We believe that driving towards an overall explanation is a fundamental characteristic of science, and therefore, it is worthwhile to have students do the same. While the current study was with adolescents learning projectile motion, studies with college students have found similar results in other domains (e.g., Schwartz & Bransford, 1998; Hallinen, et al., 2012; Shemwell, Chase, & Schwartz, 2015). The studies described here also show that despite the best of intentions, instruction can backfire. An activity designed to achieve one outcome leads to exactly the opposite. A classic example involves the hidden cost of reward (Lepper, Greene , & Nisbett, 1973). Educators tell students that they will be rewarded for doing an activity, which can demotivate an otherwise motivated student for that very activity later on. Similarly, providing people with concrete examples to make general principles more meaningful can inadvertently yield situation-specific rather than principle-based reasoning without careful implementation (Goldstone & Son, 2005). The results presented here indicate that the familiar activity of compare and contrast has a tendency to lead students to catalog features along familiar dimensions, even though the goal is for them to find relations among features. In contrast, asking students to invent an explanation for the similarities and differences appears to be a more productive task, helping students find deep structures that we want them to learn.

REFERENCES

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, 48(2), 87-113. DOI: 10.1080/00461520.2013.775712

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 640–645.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom (ASHE–ERIC Higher Education Rep. No. 1).* Washington, DC: The George Washington University, School of Education and Human Development.

Bransford, J. D., Franks, J. J., Vye, N. J. & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 470-497). New York, NY: Cambridge.

Bransford JD, Brown AL, Cocking RR. (1999). Learning from speculation to science. In: *How People Learn: Mind, Brain, Experience and School*. Washington, DC: National Research Council.

Chi, M.T.H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73-105. DOI: 10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development, 11*(3), 356-373. DOI: 10.1080/15248371003700015

Eisner, E.W. (1972). *Educating Artistic Vision.* New York: The Macmillan Company.

Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter?. In S.L. Christenson, A.L. Reschly, & C. Wylie (Eds.). *Handbook of Research on Student Engagement* (pp. 97-131). NY: Springer.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59-109. DOI: 10.3102/00346543074001059

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M.

P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. DOI: 10.1073/pnas.1319030111

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning**.** *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234-257.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*, 393-408. DOI: 10.1037/0022-0663.95.2.393

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment. *Psychological Review*, 62, 32-51

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1-38.

Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology*, *46*(4), 539-550.

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, *14*(1), 69-110. DOI: 10.1207/s15327809jls1401_4

Hallinen, N.R., Chi, M., Chin, D.B., Prempeh, J., Blair, K.P., & Schwartz, D.L. (2012, August). Applying cognitive developmental psychology to middle school physics learning: The rule assessment method. *Physics Education Research Conference Proceedings*, Philadelphia, PA.

Hatano G, Inagaki K. (1986). Two courses of expertise. In: H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child Development and Education in Japan*. New York: Freeman.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141-158.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology, 28(1)*, 129-137.

Maloney, D. P. (1988). Novice rules for projectile motion. *Science Education, 72*(4), 501-513.

Marton, F., & Booth, S. (1997). *Learning and Awareness*. Mahwah, NJ: Erlbaum.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223-231. DOI: 10.1002/j.2168-9830.2004.tb00809.x

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 177-192. DOI: 10.1002/wcs.1336.

Rittle-Johnson, B, & Star, J. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge: An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*, 561-574. DOI: 10.1037/0022-0663.99.3.561

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, *101*(3), 529-544. DOI: 10.1037/a0014224

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition & Instruction*, 16, 475-522.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology,* 103(4): 759-775. doi:10.1037/a0025140.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129-184.

Shemwell, J., Chase, C., & Schwartz, D. L. (2015). Seeking the general explanation: A test of inductive activities for learning and transfer. *Journal of Research in Science Teaching, 52(1),* 58-83.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science, 34,* 776-806. DOI: 10.1111/j.1551-6709.2010.01113.x

TABLES

Table 1: Transcriptions and Coding of Sample Worksheet Answers. Worksheets were coded by the number of physics quantities that were put into a relation in each statement.

Actual Statements	Physics Concepts	Code
They are different companies	-	Non-Factor
They are all cannons	-	Non-Factor
They start at different height	Height	Single-Factor
They go different speeds	Speed	Single-Factor
more speed = further Distance	Speed-Distance	Double-Factor
Higher drop = faster speed	Height-Speed	Double-Factor
spd × hgtm = r [key: spd=speed, hgtm = hang time r = range]	Speed-Time-Range	Triple-Factor
d = rt (distance = rate X time) To find where the person will land, you multiple the hang time by how many meters per second.	Distance-rate-time	Triple-Factor

Table 2. Percent of groups generating single-, double-, and triple-factor statements on their Phase 2 contrasting cases worksheet, Study 1. Due to the odd number of students, the 21 CC students were split into 9 pairs and one group of three students. Sixteen Invent students were split into 8 pairs. All CC groups generated at least one single-factor statement, half were able to generate a double-factor statement, but only one group was able to come up with a triple-factor statement. All Invent groups were successful in generating a triple-factor statement.

	Single-Factor	Double-Factor	Triple-Factor
Compare & Contrast (CC) n = 10 student groups	100%***	50%*	10%***
Invent n = 8 student groups	0%	0%	100%

Note: By Fischer's Exact Test for each type of factor, *** $p < .001$, * $p < .05$.

FIGURE CAPTIONS

Figure 1. Prescriptions for the selection of cases.

Figure 2. Examples of cases and a task designed to support the noticing of underlying functional relations among variables. A) This worksheet shows six different projectile motion cases, each shot at an angle of 0° from the horizontal. There are two cases for each of three companies, which each use a different starting height. Cases exhibit the effects of initial speed and height. The cases also included controlled variation: Shots by a given company varied speed but started at the same height; across companies, there are cases that share the same initial speed but have different starting heights. B) The cover page provides the whimsical context of an amusement park which shoots visitors out of cannons. The park has hired the students to figure out where to place the pool so that each visitor gets a "good landing." The example at the bottom defines the variables for each shot, including time, speed, range, and height of the cannon.

Figure 3. Interface of the PhET Projectile Motion simulation. (http://phet.colorado.edu/en/simulation/projectile-motion). Learners can control relevant variables for each shot and aim for the targets. They press "Fire" to see the projectile path, as well as the current distance, height, and time as the projectile moves through the air. The black "+" signs on the trajectory indicate 1 second intervals. Here the simulation shows a horizontal (0°) shot to simplify the domain for middle school students.

Figure 4: General procedure of research studies. The experiments had three main phases of instruction and two major assessment points. In Phase 1, students worked individually to familiarize themselves with the simulation and projectile motion. In Phase 2, students worked in pairs, using the cases shown in Figure 2. The only treatment difference occurred during this phase. In the service of placing the water tank properly for each visitor, groups received either the task to compare and contrast the companies or to invent a general solution. Students then completed a mid-test in which they individually solved a word problem on projectile motion. In Phase 3, students returned to the simulation to work individually on a Predict-Observe-Explain activity. Finally, they completed a post-test that used an isomorph to the mid-test problem.

Figure 5: Four cases used in the Predict-Observe-Explain (POE) activity. Students had to draw and predict each of the four shots, and explain how they came up with their predictions. They were then asked to use the simulation to test their predictions, and record the actual ranges. A final reflection piece asked students to explain how their predictions, if needed, could be improved.

Figure 6: Example of worksheet statements from one student group in the CC condition. Students received credit for each *type* of concept or relation they used. In this example, students only received one credit for the factor of "speed" when they noted the "same speed " in their

comparisons between the Kiddie Cannons and Blast-o-rama shots (10 m/s), and between Safe & Sane Cannons and Blast-O-Rama (8 m/s). They also received one single-factor credit for noting the range difference between the company shots, and another single-factor credit for noting the different heights.

Figure 7. Test performance on the word problem, Study 1. A) At Mid-test, after completing their contrasting cases worksheets in Phase 2, the Invent students significantly outperformed their CC peers on the quantitative word problem ($p = .003$). B) Immediately after the POE activity in Phase 3, the Invent condition maintained its advantage on an isomorphic post-test problem ($p = .001$).

Figure 8. Performance on word problem, Study 2, by condition and triple-factor generation. Inventing leads students to triple-factor generation more often than Comparing & Contrasting (100% vs. 53%). Finding the functional relation, regardless of condition, leads to better test performance than not finding the relation (85% correct vs. 25% correct, respectively).

Figure 1. Prescriptions for the selection of cases.

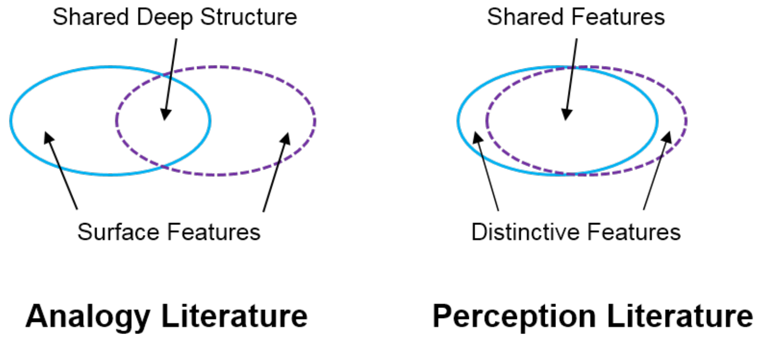

Figure 2. Examples of cases and a task designed to support the noticing of underlying functional relations among variables.

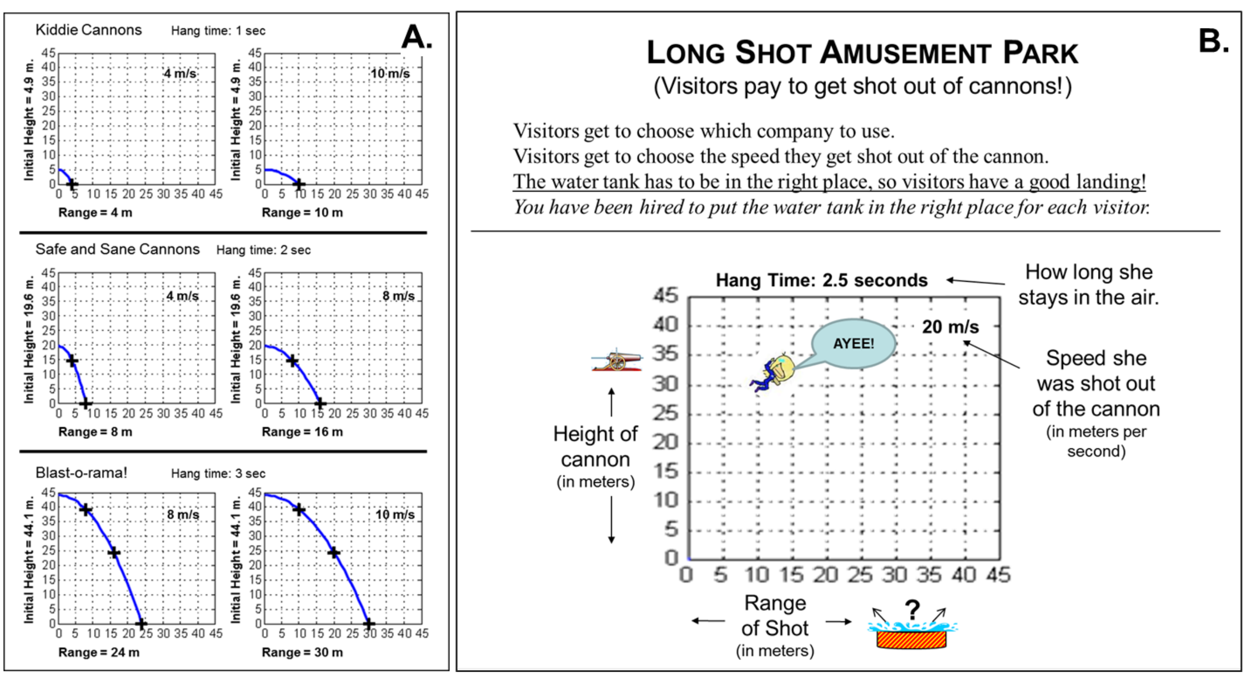

Figure 3. Interface of the PhET Projectile Motion simulation.

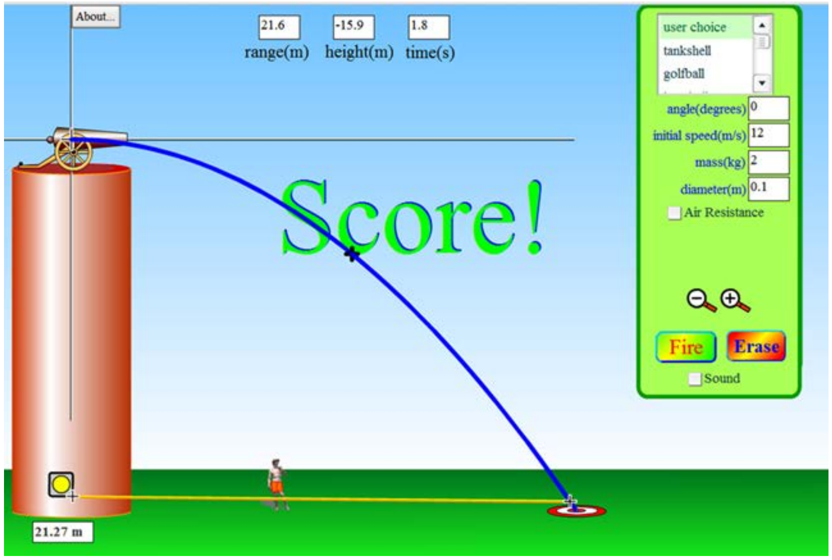

Figure 4: General procedure of research studies.

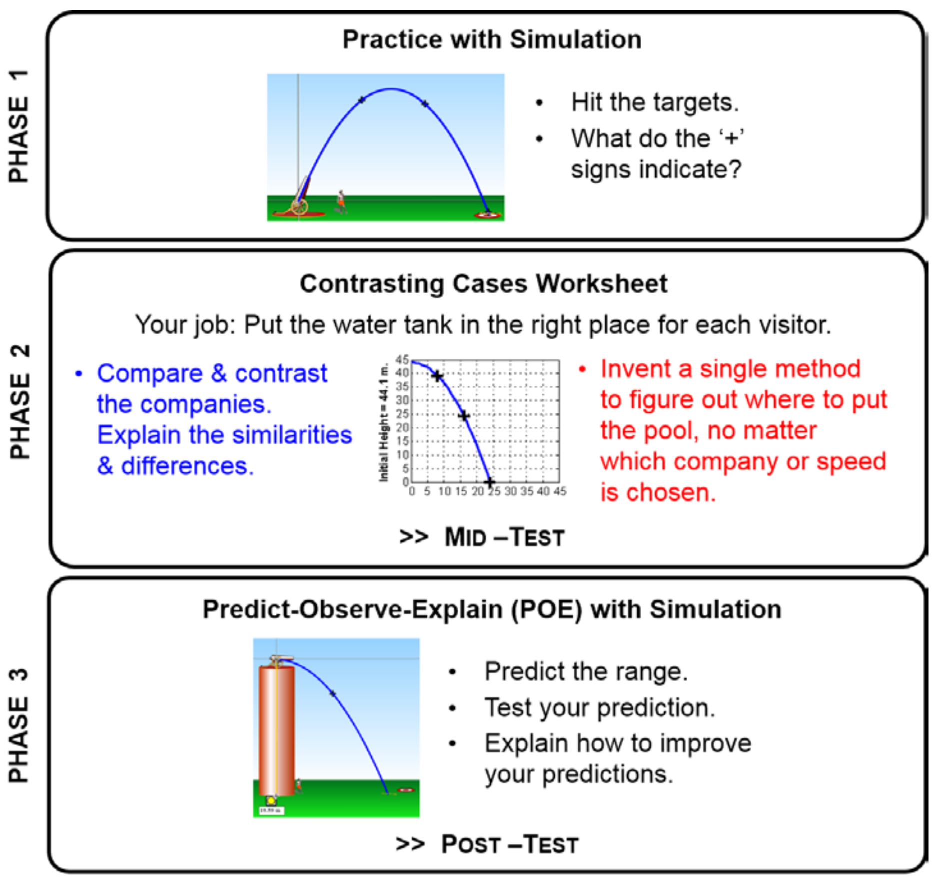

Figure 5: Four cases used in the Predict-Observe-Explain (POE) activity.

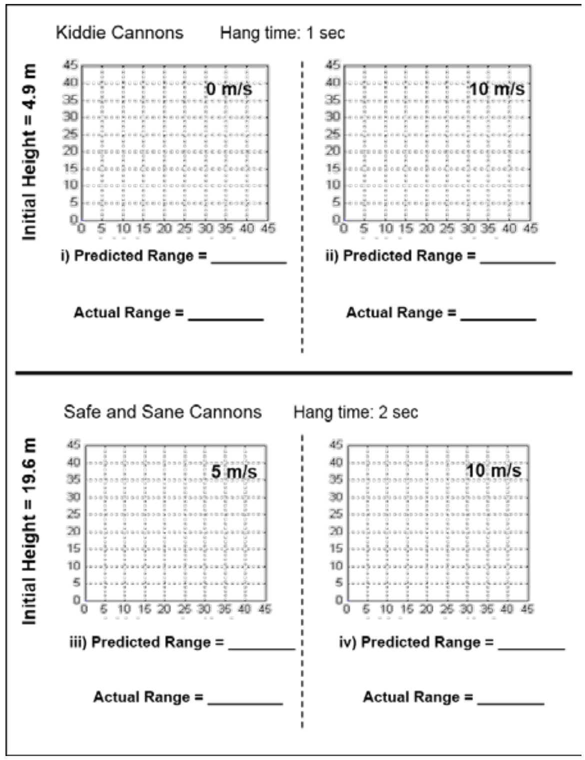

Figure 6: Example of worksheet statements from one student group in the CC condition.

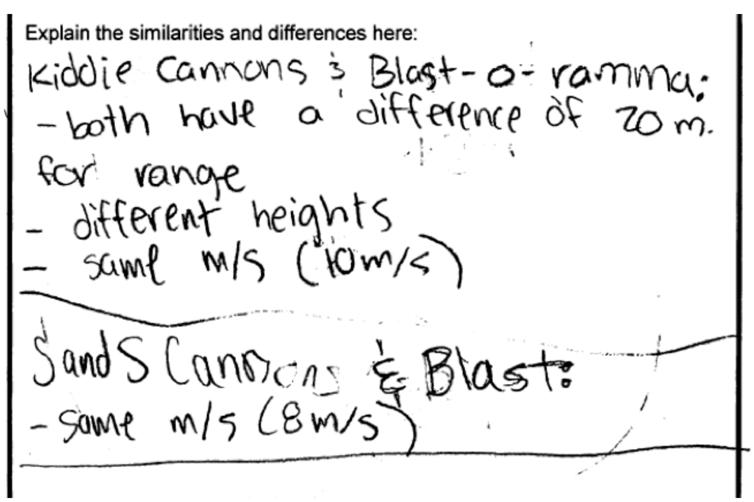

Figure 7. Test performance on the word problem, Study 1.

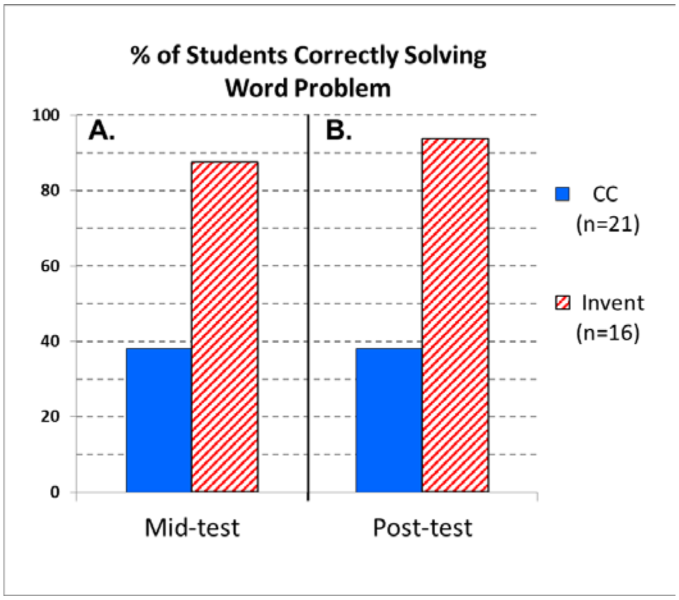

Figure 8. Performance on word problem, Study 2, by condition and triple-factor generation.