RESEARCH REPORT
Science ASSISTments Coding Scheme for Open Response Items

An Interactive Qualifying Project Report

submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

Patrick S. Sebuwufu

Date: July 22 2009

Professor Janice Gobert

Abstract

This project designs and implements a coding scheme for assessing student content knowledge and inquiry skills, as derived from their open response answers, in the domain of physical sciences. In addition, it uses the data obtained from the coding scheme to perform a statistical analysis, which is used to compare students' learning of Control for Variable Strategy (CVS, henceforth) across three different learning conditions, namely: direct instruction with reification, direct instruction without reification, and discovery learning.

Table of Contents

ABSTRACT 2
OVERVIEW 4
GOALS OF THIS PROJECT 4
INTRODUCTION 5
TEACHING INQUIRY 6
BACKGROUND 7
METHODOLOGY 9
PRETEST DESIGN PROCEDURE 11
PLANNING AND CONDUCTING AN EXPERIMENT 14
RAMP EXPERIMENT 17
ASSESSMENT OF OPEN RESPONSE QUESTIONS FROM THE RAMP EXPERIMENT DURING INTERVENTION 18
POST TEST 20
CODING SCHEME ADOPTED TO ASSESS STUDENTS' RESPONSE ON THE POST TEST 21
STATISTICAL ANALYSIS AND RESULTS 22
TABLE 4: MEANS AND STANDARD DEVIATION OF POST TEST ITEMS 24
CONCLUSION 25
REFERENCES 26
APPENDIX A: PRETEST QUESTIONS 28
PLANNING AND CONDUCTING AN INVESTIGATION 28
TIRE SWING 29
APPENDIX B: QUESTIONS DURING INTERVENTION 31
APPENDIX C: RAMP POST TEST 34
APPENDIX D: STATISTICAL ANALYSIS 35

Overview

Anyone who listens to the news today or reads the papers knows about the growing demand for improving science skills in American students. As part of that reform effort, Prof. Janice Gobert and her team managed to secure a grant from the National Science Foundation (Gobert, Heffernan, Ruiz, & Kim, 2008; NSF-DRL# 0733286) to help extend the already existing and promising Math ASSISTments program in order to address science education. The Math ASSISTments system (www.assistments.org) was designed by WPI computer science professor Neil Heffernan and his colleagues at Carnegie Mellon University. It is an intelligent tutor designed to provide students with a guided approach to answering math problems. Built into the system is the capability to create static content, i.e., multiple-choice, fill-in-the-blank, and open response questions. Within each question, images, videos, and flash images can be embedded alongside text (Gobert, Heffernan, Ruiz, & Kim, 2008).

One of the long-term goals of the ASSISTments program of research for both Math and Science is to improve students' scores on the MCAS. The system challenges students with questions from past MCAS exams and it uses a technique of providing hints, buggy messages, and scaffolds to progressively tutor a student towards the correct answer. The hints and scaffolds are provided on a just-in-time basis, i.e., when a student requests or when the system realizes the student is straggling with answering the question. Buggy messages on the other hand, are automatically generated by the system when it cross-checks the student's incorrect response with a pool of incorrect answers and then gives the student a reason why he/she might be incorrect.

The Science ASSISTments program, led by Prof. Janice Gobert and Prof. Neil Heffernan, is aimed at developing techniques to improve students' inquiry skills early in life, namely middle school, and to foster their acquisition of knowledge in an array of science domains. The initial focus of the program was placed on Physical Science.

Goals of this Project

One of the goals of this project was to determine if middle school students' knowledge about inquiry could be assessed using a test involving multiple choice and open response questions. We used data collected from a study conducted that examined which teaching approach led to better acquisition of the control of variables strategy using virtual materials (Sao Pedro et al, 2009). Students attempted to construct unconfounded experiments using a virtual ramp environment. Additionally, students answered multiple choice and short answer questions. Though ASSISTment was able to automatically score the experimental set-ups and multiple choice questions, the techniques that exist to automatically evaluate free text responses are not appropriate for these data. To get a fuller picture of students' knowledge about inquiry and, in particular, their acquisition of the control of variables strategy, it was necessary to evaluate these open-ended responses.

This project made several contributions towards analyzing the open-ended data in a deeper manner. First, we designed a coding scheme for our open response questions that matches with the National standards for tapping into students' inquiry skills in the domain of physical sciences. Second, we evaluated, using the developed scheme, students' inquiry skills in this domain. Finally, we performed statistical analyses of the results obtained from actual students' performance and proposed a better approach towards helping students improve scientific inquiry skills.

Introduction

Inquiry-based learning is rooted in the scientific method of investigating phenomenon. It is a systematic and investigative thinking capability obtained after a person has acquired a broad critical knowledge of a particular subject matter through a formal learning process (Kyle Jr, 1980). This process of learning scientific inquiry involves engaging students in the skills necessary to do inquiry. The skills include:

- Identifying questions that can be answered through scientific investigations
- Designing and conducting scientific investigations
- Using appropriate tools and techniques to gather, analyze, and interpret data
- Developing descriptions, explanations, predictions, and models using evidence
- Thinking critically and logically to make relationships between evidence and explanation
- Recognizing and analyzing alternative explanations and predictions
- Communicating scientific procedures and explanations

Listed above are the skills the inquiry skills that the NSES (National Committee on Science Education Standards and Assessment, 1996) requires students to learn by the time they finish 8th grade. Deep understanding of scientific inquiry will be reflected in students' skills to do the following:

- Skills necessary to suggest different investigation approaches, e.g. description of organisms, objects and events; discovery of new objects and phenomena, etc.

- Use mathematics and technology to analyze and quantify results of investigations
- Advance science through legitimate skepticism
- Come up with new ideas and phenomena for study

All the skills listed are neither easy nor inexpensive to assess. Additionally, standardized exams like the MCAS do very little to assess the acquisition of these skills because the exams prioritize rote learning of knowledge. As a result, teaching inquiry is often replaced with rote learning of facts and formulas to solve scientific problems. Another underlying factor limiting the implementation of inquiry is that for it to be learned, it requires authentic assessment directed towards a specific skill (Gobert, Heffernan, Ruiz, & Kim, 2008).

Teaching Inquiry

There is a mixed debate among researchers about the efficacy of inquiry-based learning. The central issue of the debate is the suggestion that inquiry learning is minimally guided so a better approach would be direct instruction as a teaching strategy. Direct instruction involves providing the information that fully explains the concept and procedures that learners are required to know (Kirschner, Sweller, & Clark, 2006). Discovery learning on the other hand, is a learning context in which instead of providing all the necessary information, learners must discover or construct the information themselves (Bruner, 1961; Papert, 1980; Steffe & Gale, 1995).

Opponents of the discovery learning strategy argue that problem-solving skills are derived by drawing from existing experiences stored in long-term memory and this memory is shaped by direct instruction (Kirschner et al, 2006). They go further to pinpoint that the capacity of working memory is limited to 7 elements (Peterson & Peterson, 1959) and its duration to 30 seconds during which information can be kept before it is lost (Miller, 1956). Most importantly, Kirschner et al stress that in the interaction between long-term and working memory, capacity and time limits only apply to information not yet stored in long-term memory. They add that inquiry-based instruction is inefficient because it places a large burden on working memory (Sweller, 1988). Other critiques suggest that when inquiry tasks are too open-ended, students become lost and frustrated and their confusion can lead to misconceptions (Brown & Campione, 1994), so as a result, teachers spend considerable time scaffolding students on procedural skills (Aulls, 2002) making it difficult to tailor students to individual needs in real time within a classroom setting (Fadel, Pasnick, & Pasnick, 2007). Additionally, during inquiry learning students can have difficulties designing experiments, forming a testable hypothesis, and drawing correct conclusions

(Klahr & Dunbar, 1988; Kuhn, Garcia-Mila, Zohar & Andersen, 1995). Furthermore, these difficulties can hinder the learning of targeted scientific principles (Kirschner et al, 2006).

On the other hand, supporters of inquiry-based instruction clarify that the strategy are typically *not* minimally guided (Hmelo-Silver et al, 2007) because it is full of scaffolding procedures which use direct instruction (Krajcik, Czerniark & Berger, 1999; Schmidt, 1983; Schwartz & Branford, 1998) provided on a just-in-time basis, i.e., when the learner needs to know the information (Edelson, 2001) and that this, in turn, promotes learning. Hmelo-Silver et al, additionally point out that scaffolding inquiry tasks encourages learners to engage in complex tasks that would otherwise be beyond their current skill level (Rogoff, 1990, Vygotsky, 1978; Quintana et al, 2004; Hmelo-Silver et al, 2007). In addition to that, other studies suggest that children who acquire knowledge on their own are more likely to extend that knowledge than those who receive direct instruction (Bredderman, 1983; McDaniel & Schauble, 1996; Storhr Hunt, 1996). Other researchers have shown that students can effectively learn science inquiry strategies. For example, Klahr and Niagram (2004) focused on the control for variable strategy (CVS). In their study students were placed in one of two learning conditions: discovery learning or direct instruction. In the discovery condition students used an actual ramp apparatus to construct unconfounded experiments to determine which factors led a ball to run the furthest down a ramp. In the direct instruction condition, several ramp apparatus were set up for students and they were asked to determine which experiments were confounded in a think-out-loud protocol. Although the students in the direct learning condition outperformed those in the discovery learning condition, the study helped suggest a new mechanism called "path independent learning" which means that if learning occurs, performance will be the same irrespective of learning condition (Klahr & Nigam, 2004).

Background

Klahr and Nigram (2004) conducted their investigation to evaluate the relative effectiveness of direct instruction and discovery learning with respect to acquisition of control for variable strategy (CVS). They also wanted to test the earlier findings "that what is learned is more important than how it is taught." To do this they used the ramp apparatus shown in the Figure 1 below.

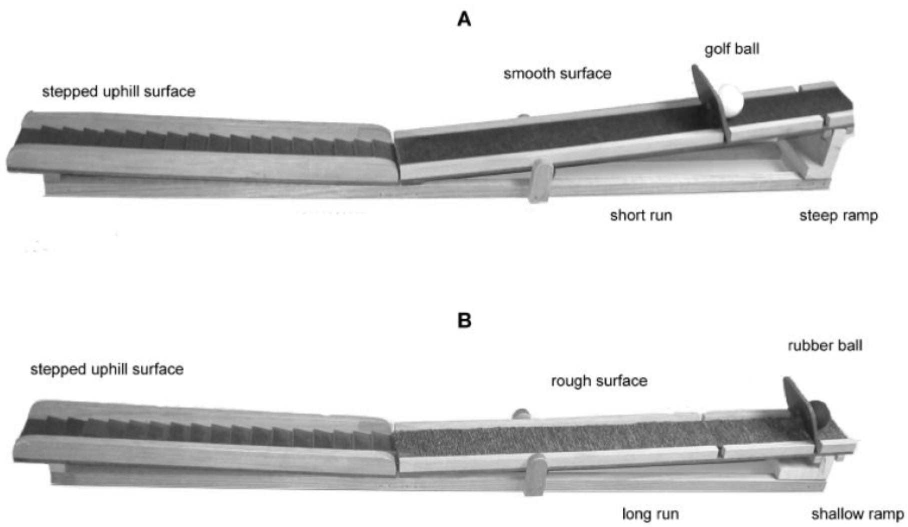


Figure 1: Ramp Apparatus used in the Experiment (Klahr & Niagram, 2004)

The experiment used a wooden ramp apparatus that had four different variables that could be controlled, namely:

- Steepness (High or Low)
- Run Length (Short or Long)
- Surface (Rough or Smooth)
- Ball type (Rubber or Golf)

The students were asked to make comparisons that would determine how different variables affected the distance the ball rolled off the ramp. In the direct condition, the instructor provided good and bad examples of CVS; explained what differences were between them and also told students how and why CVS works. The students were then asked to determine which experiments were confounded. In the discovery learning condition on the other hand, there were no explanations, examples or feedback from their performance.

During assessment the students were to set up different ramp settings, four ramp settings altogether: Two of them would determine the effect of a factor (run length) that had been investigated earlier and the other two determine the effect of a factor (surface) that had not been covered in the investigation. The students' content knowledge on CVS was measured based on how well they could evaluate science fair posters about two weeks after the first investigation.

The results from the experiment showed that students in the direct instruction condition outperformed those in the discovery condition in both the assessment and near transfer task (Klahr and Nigam, 2004).

The present project utilizes data captured as part of an experiment that was an extension of Klahr and Nigam's (2004) study on CVS conditions (Sao Pedro et al, 2009). The difference was in the approach taken, i.e. instruction was to be given using a virtual ramp environment. Additionally, the extension of the study added another learning condition, i.e., the direct-without-reification condition. The ramp environment was created using the ASSISTments system. The system has the built in capability of automatically assigning students to different learning conditions in the same classroom; this feature was utilized in the present study.

Methodology

Pretest

The pretest was composed of questions designed to assess students' inquiry skills. The pretest questions were specifically designed to assess student skills in hypothesizing, defending their hypothesis, planning and conducting an experiment, and also familiarizing themselves with the ramp apparatus.
The purpose of the pretest was to find out how well students were versed with inquiry applications.

Ramp Experiment

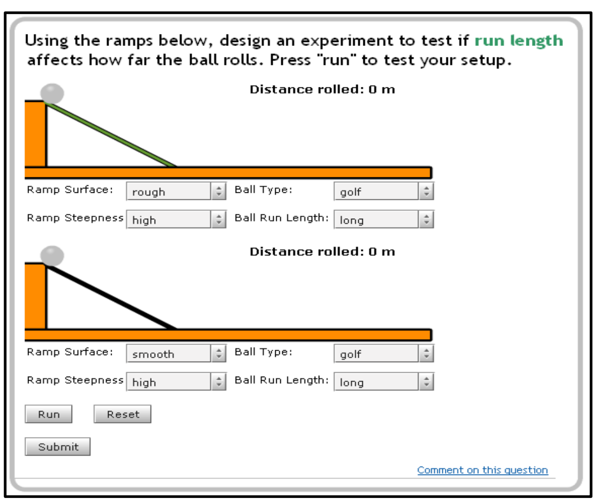


Figure 2: Sample Ramp Setup (Gobert, Heffernan, Ruiz, & Kim, 2008)

The ramp environment shown in Figure 2 was developed using Open Laszlo framework (www.openlaszlo.org). The environment enables development of a rich internet application (RIAs) that can easily compile out to Flash format which is compatible with ASSISTments. The virtual ramp apparatus in comparison with Klahr's experiment also had four variables, which could be manipulated: Steepness (high or low), Run length (short or long), Surface (rough or smooth), Ball type (rubber or golf) The students' task was to setup, predict, and run the experiment depending on the question brought forward to them. On the ramp setup pressing the "run" button will cause the ball to roll down the ramp and the distance the ball rolls will be displayed in meters by the scale. The "Reset" button will cause the ball to be placed back on the ramp and the "submit" button will log the student's data.

The system could automatically place students in the same class into three different learning conditions namely: Discovery learning, direct-with-reify and direct-without-reify. The ordering of the questions brought forward to students is summarized in Figure 3 below.

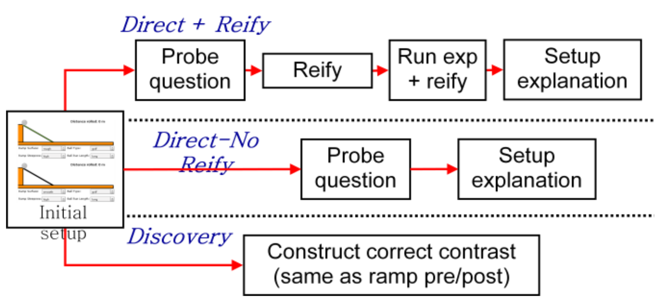


Figure 3: Experimental Design (Gobert, Heffernan, Ruiz, & Kim, 2008).

Students were allowed to manipulate the ramp setup while changing the variables as many times as they wanted until they came up with a satisfactory result that would be submitted and logged by the ASSISTments system.

Development

In addition to solving interactive inquiry problems involving the ramp, during the pretest, intervention, and posttest students were tested with multiple choice, fill-in-the-blank, and open response questions. The multiple choice and fill-in-the-blank questions were automatically scored by the ASSISTments system. However, the system cannot presently interpret written language, so someone needed to directly code the responses from the students to characterize their inquiry skills.

The ramp experiment was conducted by students during class time over three class periods and on three consecutive days with each period being sixty minutes long. On the first day, students took an inquiry pre-assessment and a survey. In the inquiry pre-assessment, the students answered standardized multiple choice and open response questions with three of the multiple choice questions pertaining to CVS. On the second day, the students were introduced to the ramp environment and its variables, the ramp pretest, intervention and posttest, which tested for both inquiry and content knowledge. During the ramp pretest, all students regardless of learning conditions demonstrated their understanding of CVS by constructing unconfounded setups to test if a particular variable affected how far the ball rolled.

During the intervention, students were presented with ramp setups, some confounded and others unconfounded, irrespective of the learning condition. However, the nature of interacting with the ramp environment varied depending on the learning condition. In the direct + reify condition, students were given an overview of CVS before answering the question. After, they were given a “Yes/No” response and an open response question to justify their answer and, they were given an explanation why the experiment was confounded or unconfounded. The direct-no-reify condition followed the same procedure except students were not asked an open response question. The discovery learning condition was exactly similar to the other two conditions except students were not given any information regarding CVS or even input on the correctness of their answer (Sao Pedro et al, 2009).

Pretest Design Procedure

The pretest was designed to both align closely with and elicit the scientific inquiry skills identified in national science standards (NSES, 1996). Specifically, students were engaged in the following tasks:

- *Hypothesizing* about the likelihood that watering seeds would affect how they germinate; how length of a string affects how far a pendulum can swing
- *Observing* experimental data collected and develop logical reasoning to defend their hypothesis
- *Planning and conduct an experiment* to test the hypothesis they just made

- *Analyzing* data obtained from an experiment and determine which variables affected specific aspects of the experiment, e.g., which variable made the tire swing the fastest in the tire-swing experiment.

Hypothesis Question

On the hypothesis question, students were offered an example experiment previously performed and tested by a hypothetical student who wanted to hypothesize about whether covering lettuce seeds with soil would help the seeds to germinate not. The experimental procedures and table of result were shown to the students. Based on the experimental procedure and results provided, the students were requested to hypothesize on an extension of the experiment they had seen, i.e., they were asked to make a hypothesis which tested if watering seeds daily with 100ml of water would affect germination. The question is shown in Appendix A.

In this question, we wanted to see how much inquiry knowledge students possessed with regard to hypothesizing. We were specifically interested to see if the students could provide a rich scientific hypothesis containing aspects such as:

- Description of the observation brought forward to them
- A clear, logically coherent description of the process hypothesized to cause the observation
- How well they understood and also employ variable control strategy to scientific experiments
- Description with consistency of the above aspects supported by facts and accepted theories

Assessment of Students Response on the Hypothesis Question

Using the national inquiry standards for accessing inquiry (NSES 1996), for each of the aspects listed above in the design section, specific rubrics were developed to score students' response. Coding was done on a two point scale (0,1) or (yes, no) and other times on a three point scale (0,2), depending on how much information the specific item carried in regards to the inquiry task we were looking for in a student's response. A detailed explanation of the coding scheme developed for each of the items listed above is provided below.

Item Pre1A: Student's motivation and also understanding scientific question

This facet did not play a large role in the larger project, i.e., extending Klahr and Nigram's experiment on variable control strategy, therefore, it was coded on a two point scale and it did not have subsections.

Student provided a meaningful response.	Yes
Student left question blank, just scribbled or gave an answer suggesting he/she did not understand the question.	No

Item Pre1B: control for variable strategy

This item carried the greatest load with regard to the larger project, so specific emphasis was placed on it. It was divided into subsections to provide a finer grade lens with regard to students' understanding of the subject.

(a) **Dependent Variable**

In this section, we were only interested in finding out if the student knew about the dependent variable. Correctness did not matter in this subsection. All that was required was for the student to mention the dependent variable or provide a statement that indicated that the student is aware of the variable.

Student mentions a dependent variable, or his response indicates he recognized presence of a dependent variable.	1
Student does not mention presence of a dependent variable or his response does not indicate he recognized presence of one	0

(b) **Independent variable**

This item was coded in exactly the same manner as the subsection above except this time the students were tested for the independent variable.

(c) **Relationship between independent and dependent variable**

This item was designed to test how students relate the independent and dependent variables to each other to make sense out of the experiment.

Student mentions any relationship between the independent and dependent variable.	1
Student does not provide any relationship between the independent and dependent variable	0

(d) **Correctness of relationship between the independent and dependent variable**

For those students who realized there was a relationship between the independent and dependent variable, we were interested in finding out if the relationship the students provided was correct. This item was graded on the scale (0,1), where those who got the answer correct got a 1, and those who did not obtained a zero.

(e) **Specific relationship between independent and dependent variable**

This section was designed to make an even finer grade lens on students' understanding of CVS. If students provided a correct relationship between the independent and dependent variable, this section sought to find if those students that provide the correct relationship also realized the trend the two variables follow, i.e., a positive or negative trend. This subsection was also graded on a (0,1) scale.

Planning and Conducting an Experiment

Students were given a detailed explanation of what a pendulum is and how it works. Then a sample pendulum was constructed out of a coffee can and a picture of it was shown to the students. The students were then expected to provide a testable hypothesis explaining how changing the length of the string would affect how far upwards the pendulum would swing. Additionally, the students were required to describe how they would conduct and test the hypothesis they just made. This section was designed to reinforce the students' hypothesizing skills while at the same time introducing a link between a hypothesis and an experiment that can actually test it. The main facets of this question included:

- Motivation and planning
- Hypothesis testing
- Organization
- Control for variable strategy

Item Pre2A, Pre2B: Motivation and realistic planning

Item Pre2A was coded in the same manner as item Pre1A above. This section also had a subsection (item Pre2B) which tested how realistic a student's experiment was, i.e., we were interested in finding out if the experiment the student provided could be performed if the student had access to the necessary equipment and time to carry out the experiment. The table below shows the coding scale for item Pre2B.

Student provided a realistic experiment that could be tested given time and necessary equipment	1
Student's experiment was unrealistic and could not be conducted even if student was given time and materials necessary	0

Item Pre2C: Hypothesis testing

This aspect was designed to check if the approach students took was directed towards testing the hypothesis they had provided or not. It was coded for correctness on a three-point scale (0,2).

Student exactly "tests" hypothesis they provided	2
Student's experiment provides some evidence of "testing" the provided hypothesis	1
Student's experiment does not "test" the hypothesis they provided	0

Organization

This facet was broken down into three different subsections each testing a different inquiry task as listed in items Pre2C, Pre2D and Pre2E below. This item was not coded for correctness, but we were just interested in finding out if the student employed the skill.

Pre2D: Collecting materials

On a two point scale, as listed in the table below, students, whose response included collecting the equipment necessary to conduct the experiment they proposed, were given credit.

Student mentions gathering the experimental materials	1
Student does not mention gathering experimental materials	0

Pre2E: proper scientific procedure

Student talks about recording and processing data they obtained from the experiment.	2
Student talks about recording data they obtain from the experiment, but they don't indicate processing it.	1
Student does not record the data they obtain from the experiment	0

Pre2F: Measuring experimental outcomes

Student mentions measuring experimental results	1
Student does not measure experimental results	0

Pre2G: Control for variable strategy

This facet was a major goal of the larger project. Evidence of applying CVS was explicitly coded for in every question brought forward to the students.

Student only changed the target variable while holding all the other variables in the experiment constant	2
Student mentions changing the target variable they mentioned in their hypothesis	1
Student does not mention changing the target variable mentioned in the hypothesis.	0

Drawing a logical Conclusion from provided data

To extend knowledge about the pendulum experiment, students were offered an example application of a pendulum. The students were provided data obtained from experiments performed by two separate students who built tire swings in their back yards. The experimenters could change the mass of the tire and the length of the string and then record the time it took for the tire to swing back and forth. The students were provided with a diagram of the tire swings and also the experimental results obtained by the experimenters in a table. They were then challenged to deduce from the data obtained from the experiments, what could be done to make the strings move the fastest. In addition to extending the students' knowledge about pendulums and their real life applications, this question went farther to test how well students analyzed scientific data to make sense out of it. Similar to the previous pretest questions, this question was coded for the following inquiry strands:

- Motivation
- Analysis

Pre3A: Motivation

The motivation facet was coded on a two point scale (Yes, No) as noted in the two motivation facets in the previous questions, i.e. "Yes" if the student's response was meaningful or "No" if the student provided a meaningless response or if they just left the question blank.

Analysis

The analysis aspect was divided into four separate subsections to provide a finer grade lens into the students' inquiry knowledge. The question was coded partly for correctness but we were also looking for students providing a response, which was supported by the data provided to them.

Pre3B: Basing response on data provided

Student mentions using the table provided	1
Student does not mention using the table	0

Pre3C: Specific details from provided information

Though this aspect is very similar to item Pre3B, we were interested in finding out if students are comfortable with providing factual information, which is supported by the data given to them in order to derive logical conclusions.

Student used specific details from the table	1
Student does not use specific details from the table	0

Pre3D: Correctness of student's conclusion

This facet was meant to find out how well the student could draw conclusions based on the data that was provided to them, i.e., we wanted for students to look at the table and be able to deduce that the shorter the rope was, the faster the pendulum would move.

Student provides a correct conclusion based on the table that was provided	1
Student's conclusion is wrong or the student response does not indicate that the results are based on the provided information	0

Pre3E: Detailed analysis

This facet pointed out those students who could look at the table and then realize that the mass of the tire did not contribute anything to how fast the tire would swing. This indicated that the students deeply understood the experiment.

Student drew a correct relationship between mass of the tire and speed of the pendulum	1
Student did not realize the relationship or drew the wrong conclusion	0

Ramp experiment

The ramp shown in Figure 2 was used to test for the student's understanding of CVS as emulated in the Klahr and Niagram's study. Students were allowed to manipulate the ramp apparatus so they could learn how to use it. A pretest was offered to see how well the students had learned in regards to using the ramp setup. The pretest was only composed of multiple choice questions and it was directly graded by the system. For that reason, the following sections will describe what happened during the intervention and the post test on the ramp apparatus.

The type of intervention offered to the students depended upon the learning condition they were placed in. During the session, students were required to set up two ramps so that the target variable was contrasted while the other variables were held constant. The ramp had four different states, i.e., *unconfounded* (all variables were controlled), *singly confounded* (one variable was controlled), *multiply confounded* (more than one variable was controlled), and *uncontrasted* (the target variable was unchanged).

Students were then asked a series of ASSISTments questions related to the interactive plane environment. Each of the questions required students to perform some combination of reading, descriptions, designing, and running experiments. The questions were composed of multiple choice, fill-in-the-blank, and open response questions. The students were required to answer a question predicting the outcome of each of the experiments before they ran it. The same question was asked after each experiment was run to see if the students had changed their minds along the way. The questions asked during the intervention are shown in Appendix B.

Assessment of Open Response Questions from the Ramp Experiment during Intervention

With regard to inquiry, the open response questions on the ramp experiment were coded no differently from the pretest. Additionally, the coding scheme used was identical irrespective of the learning condition. The scale used to assess students' responses only differed in magnitude in some sections depending on the number of variables that were controlled. The inquiry aspects that the coding scheme intended to capture included:

- Motivation
- Consistency throughout the experiment
- Correctness of prediction
- CVS understanding and application
- Recognizing confounded variables

Since the questions asked before and after the students ran the experiment were similar, the same coding scheme was adopted for both the prediction and the response provided after the students ran the experiment.

Ramp1A: Motivation

The motivation facet was coded in exactly the same manner as it was in the pretest, i.e., on a two point scale (Yes/No) with "Yes" representing a meaningful response and "No" representing a meaningless response or a question that was left blank.

Ramp1B: Consistency throughout the experiment

In this facet we checked to see if the open response provided by the student was consistent with the "Yes/No" response they provided in the multiple choice answer they gave. Also, since the same coding scheme was adopted to grade the open response the students provided after they ran the experiment, this facet tested to see if the two responses the students gave did not change in meaning.

-Student response is consistent with (Y/N) response in the previous question. -After running the experiment, the response the student gives is consistent with the one provided before	Yes
-Student's response is not consistent with the (Y/N) response they gave in the previous part -After running the experiment, the response provided is not consistent with previous response	No

Ramp1C: Correctness of prediction

Student's statement is true given the ramp setup	1
Student's statement is incorrect	0

CVS understanding and application

This aspect was divided into subsections to get a better understanding of the students' knowledge of CVS since it is the major goal of the larger project. It was coded for both correctness and also evidence of knowledge with regard to the control for variable strategy. The coding scheme for these facets is reflected in items Ramp1D-F.

Ramp1D: Evidence of variable control strategy

This facet was not graded for correctness but on a two-point scale students were credited for providing statements that showed they had a crude idea of the technique.

Student's statement reveals some CVS knowledge	1
Student's statement does not reveal any CVS knowledge	0

Ramp1E: Identification of dependent variable

Student correctly identified the dependent variable	1
Student did not correctly identify the dependent variable	0

Ramp1F: Correctness of relationship between target and other variables

Student correctly relates the dependent variable to independent variable(s) in the experiment	1
Student does not relate the dependent variable to independent variable(s) in the experiment	0

Ramp1G: Identifying confounded variables

If the variables were confounded, we tested to see if the student correctly identified which variables were confounded. If the variables were not confounded, we wanted to see if the students realized that all other variables except the dependent were controlled. In this section, for each confounded variable the student identified, they got points chosen on the scale of 0-4 depending on the number of confounded variables in a particular experiment.

Post test

The post test was composed of three questions that were meant to find out how much students had learned from the ramp experiments. The first question asked students to relate to the ramp experiment they had setup during the intervention session and then explain how they can determine if one particular variable affects the distance the ball rolls. The next question sought to find how students could relate to experiments with more than one confounded variable; the question asked students that if there was more than one variable that could be changed, to determine how each variable affected how far the ball rolled. The last question required students to look back at the experiments they performed and provide a combination of variables that would cause the ball to roll both the furthest and shortest distance. The specific questions on the post test are shown in appendix C. All three questions were scored using the same coding schemes to access students' transfer of knowledge across different inquiry domains.

Coding Scheme Adopted to Assess Students' Response on the Post Test

Post1A: Motivation

This facet was coded in exactly the same manner and scale as all the motivation facets in both the pretest and intervention sessions.

Student provided a meaningful response.	Yes
Student left question blank, scribbled, or gave an answer suggesting he/she did not understand the question.	No

Post1B: Correctness of Variable Combination

On this item, one point was earned for each variable the student correctly identified with regard to rolling the ball the farthest or the shortest distance.

An explanation offering the correct effect a variable has on the distance rolled by the ball down the ramp. A point was earned for each variable explained correctly	1 – 4
Student does not offer an explanation in relation to how any variable affects the distance the ball rolled	0

Post1C: consistency

Student's explanation is backed by data provided in the experiment	1
Student does not base his/her explanation on the data provided	0

Post1D: Control for variable strategy

Student conclusively employed variable control strategy	2
Student's response has some evidence of use of variable control strategy	1
Student did not provide any evidence of employing variable control strategy	0

Post1E: Prior knowledge claims

This facet was meant to check if the student used the inquiry knowledge they had learned from the intervention session while answering the questions from the post test.

Student’s response is entirely dependent on prior knowledge	2
Student response reflects some prior knowledge	1
Student response does not show any prior knowledge.	0

Statistical analysis and Results

The major study of which this project is a part focused on finding out which condition(s) yielded better learning results on the posttest as measured by multiple choice items (Sao Pedro et al, 2009). In particular, the study sought to address any differences across the learning conditions i.e., direct-with-reify, direct-no-reify and versus discovery learning conditions. Using the data obtained from the multiple choice questions the initial statistical analysis was conducted.

A multivariate analysis of variance (MANOVA; Pedhazur, 1982) was carried out to be sure that there were no statistical differences across the three groups on the pretest. This was confirmed; thus any group differences in the post-test could be attributed to the learning condition rather than differences in prior knowledge across the three groups going into the experiment. The study went further to find out if students learned more in any of the three conditions. To determine which condition led to improved performance on CVS inquiry and posttest scores, a multivariate analysis of covariates (MANCOVA) was done. The results showed that the dependent variable was significantly affected by the condition. Univariate ANOVA was then computed for each dependent variable (inquiry posttest and ramp posttest). The means and standard deviations are shown in Table 1 and the ANOVA results are shown in Table 2.

From the results it was noted that the effects of the condition were marginally significant for the ramp posttest and insignificant for the inquiry posttest. So because of this marginal difference across conditions on the ramp posttest score, further investigation was done to find out if certain posttest items yielded different learning gains across the conditions. The investigation tested whether performance was different for questions involving unconfounded, singly confounded, or multiply confounded items; with the multiply confounded items considered more complex than the singly confounded items. Another MANCOVA by item complexity was computed.

The results from this computation showed that the condition again was significant with respect to the combined dependents, i.e., unconfounded, singly confounded and multiply confounded posttest items.

A univariate ANOVA was then computed for each posttest dependent variable. The results showed that for the singly confounded items, condition was not significant. However, it was significant on the more complex items i.e., multiply confounded items. The results are shown in Table 3.

Table 1: Summary of Students Performance on Pretest and Posttest (Sao Pedro et al, 2009)

		Direct+Reify (N=45)		*Direct-No Reify* (N=42)		*Discovery* (N=43)	
	Max	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
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

Table 2: Univariate ANOVA factor and covariate significance for inquiry and ramp posttest items (Sao Pedro et al, 2009)

Source	Post test DVs	df	F^a	p	η^2
ELL	Inquiry	1	0.37	.547	.003
	Ramp	1	2.67	.105	.022
Grade	Inquiry	1	0.97	.327	.008
	Ramp	1	0.18	.672	.001
Inquiry Pretest	Inquiry	1	7.54	.007	.059
	Ramp	1	0.01	.907	<.001
Ramp Pretest	Inquiry	1	13.11	<.001	.098
	Ramp	1	48.31	<.001	.285
Condition	Inquiry	2	1.91	.152	.031
	Ramp	2	2.61	.078	.041

a: *df error* = 125

Table 3: Univariate ANOVA factor and covariate significance for singly confounded (SC) and multiply confounded (MC) ramp posttest items (Sao Pedro et al, 2009).

Source	Ramp post test DVs	*df*	F^a	*p*	η^2
ELL	Ramp MC	1	2.55	.113	.021
	Ramp SC	1	1.86	.175	.015
Grade	Ramp MC	1	1.44	.232	.012
	Ramp SC	1	0.00	.955	<.001
Pre UC	Ramp MC	1	0.24	.629	.002
	Ramp SC	1	1.40	.239	.012
Pre SC	Ramp MC	1	17.28	<.001	.126
	Ramp SC	1	11.11	.001	.085
Pre MC	Ramp MC	1	5.21	.024	.042
	Ramp SC	1	1.74	.190	.014
Condition	Ramp MC	2	6.76	.002	.101
	Ramp SC	2	0.33	.720	.005

Most importantly, the study found both the direct-with-reify and the direct-no-reify conditions outperformed the discovery learning condition. Additionally, the study found no significant difference between the direct-with-reify and direct-no-reify condition (Sao Pedro et al, 2009).

In this project, another statistical analysis similar to that done for the multiple choice questions was performed for the open response questions. The means and standard deviations are shown inTable 4 below. In the MANOVA test, the condition was used as the independent variable; the open response post-test (CVS and content) as the dependent variables and the inquiry pretest and ramp pretest as covariates.

Table 4: Means and Standard deviation of post test Items

Item	Direct with Reify (N=37)		Direct No Reify (N=29)		Discovery (N=35)	
	Mean	SD	Mean	SD	Mean	SD
Post_Open_CVS	4.4054	2.15329	3.6286	1.76711	4.8276	2.42117
Post_Open_Content	4.5135	3.23713	4.9143	3.35517	5.5517	3.01882

The results yielded no significant differences across the different learning conditions at the 0.05 level of alpha (Wilks' Lambda λ= .911, F=2.266[b], p=.064, η^2=.046). The detailed tables of results are shown in Appendix D. Since the results showed no significant differences, no univariate analysis was performed.

Conclusion

This project has: First, designed a coding scheme which matches the US National Standards (NSES 1996) to assess students' content knowledge and inquiry skills in physical sciences. Secondly, using the developed scheme, we evaluated students' content knowledge and inquiry skills in that domain and finally, we conducted some analysis of the data obtained from the coding scheme for the open-ended response data. Additionally, although the analysis of the automated data from the three conditions did yield significant learning gains favoring the direct learning condition (Sao Pedro et al, 2009), the analysis of the open response data from this study did not yield statistically different results in students' performance on the open-ended responses across the different learning conditions, namely: discovery learning, direct-with-reify, and direct-without-reify. This could be due to the nature of the items themselves, i.e., perhaps they were not sensitive enough to detect group differences. Further research is necessary to empirically determine whether this is the case or not. This is the first empirical study in a series of studies to be performed as part of this project, all with the goals of teaching scientific inquiry.

References

Aulls, M. (2002). The Contribution of Co-Occuring Forms of Classroom Discourse and academic Activities to Curriculum Events and Instruction. *Journal of Education Psychology* , 94,520-538.

Bredderman, T. (1983). Effects of activity-based elementary science on student outcomes: A qualitative synthesis . *Review of Educational Research* , 53, 499-518.

Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly(Ed), Classroom lessons: Integrating cognitive theory and classroom practice. *MIT Press* , 229-270.

Bruner, J. S. (1961). The art of discovery. *Harvard Education Review* , 31,21-32.

Edelson, D. C. (2001). Learning-for-use: A framework for integrating content and process learning in the design of inquiry activities. *Journal of Research in Science Teaching* , 38,355-358.

Fadel, C., Pasnick, S. H., & Pasnick, S. (2007). Assesment 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* .

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* , 100-105.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidence During Instruction Does Not Work: An Analyisis of the Failure of Constractivist, Discovery, Problem-Based, Experimental and inquiry-Based Teaching. *Educational Pychologist* , 76-84.

Klahr, D., & Dunbar, K. N. (1988). Assesment in the Age of innovation. *Cognitive Science* , 12,1-48.

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

Krajcik, J., Czerniak, C., & Berger, C. (1999). *Teaching children science: A project-based approach.* Boston, MA: McGraw- Hill.

Kuhn, D., Garcia-Milla, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition . *Society for Research in Child Development Monographs* .

Kyle Jr, W. C. (1980). The distinction between inquiry and scientific inquiry and why high school students should be cognizant of the distinction. *Journal of Research on Science Teaching* , 17,123-130.

McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction* , 7,129-159.

Sao Pedro, M. Gobert, J., Heffernan, N., & Beck, J. (2009). *Can an Intelligent Tutor Teach the Control of Variables Strategy for Scientific Inquiry?* To be published in the Proceedings of the Cognitive Science Society. Amsterdam, The Netherlands, July 20th-August 3.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity of processing information. *Psychological Review* , 63,81-97.

National Committee on Science Education Standards and Assessment, N. R. (1996). *National Science Education.* Washington DC.

Papert, S. (1980). *Mindstorms: Chidren, computers and powerful ideas.* New York: Basic Books.

Pedhazur, E. J. (1982). *Multiple Regression in Behavioural Research.* New York, N.Y.: CBS College Publishing.

Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology* , 58,193-198.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., & R. G. Duncan, e. a. (2004). A Scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences* , 13, 337-386.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* New York: Oxford Univercity 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.

Schauble, L. (1996). The development of scienfic reasoning in knowleg rich contexts. *Journal of Experimental Child Psychology* , 32,102-119.

Schmidt, G. H. (1983). Problem-based learning: rationale and description. *Medical Education* , 17,11-16.

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

Steffe, L. P., & Gale, (. J. (1995). *Constructivism in Education.* Hillsdale, NJ: lawrence Erlbaum Associates, Inc.

Stohr-Hunt, P. M. (1996). An analysis of frequency of hands-on experience and science achievement . *Journal of Research in Science Teaching* , 33,101-109.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science* , 63,81-97.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Appendix A: Pretest Questions

Hypothesis question

A student hypothesized that lettuce seeds would not germinate (begin to grow) unless they were covered with soil. The student planted 10 lettuce seeds under a layer of soil and scattered 10 lettuce seeds on top of the soil. The data collected is shown in the table below.

Seed treatment	Number of seeds germinated (beginning to grow)
Planted under soil	9
Scattered on top of soil	3

What can be concluded from this data?

Multiple choice:

- ✗ Seeds grow better if they are scattered on top of soil.
- ✓ Seeds grow better if they are planted under soil.
- ✗ We cannot be sure from this data.
- ✗ Seeds grow well in both conditions.

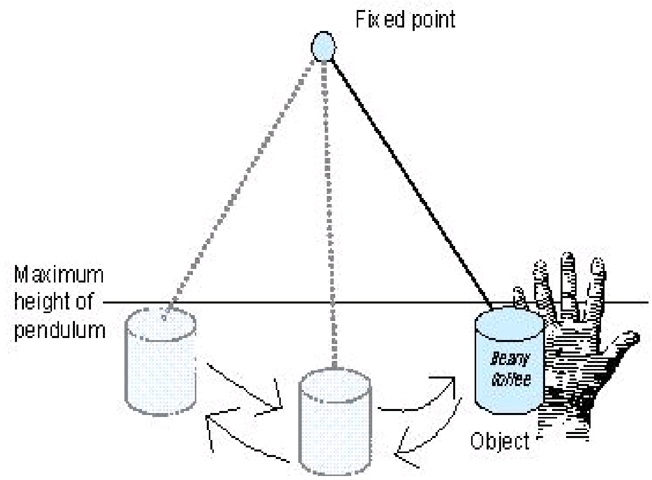
Ungraded open response:

b) Now this student wants to test whether watering the seeds daily with 100 milliliters of water will affect how they germinate. Write a **hypothesis** you could test that states how watering the seeds affect germination.

Planning and Conducting an Investigation

A **pendulum** is an object that hangs by a string attached to a fixed point. A pendulum can swing freely back and forth as it hangs from the string. The figure below shows an example of a coffee can that is made into a pendulum by attaching it to a fixed point using a string.

When the coffee can is released from its starting place, it will swing down in the direction shown by the solid arrows. When it reaches the bottom it will swing back up and reach a maximum height. Then it will stop for a small moment and begin to swing back towards its starting place. The coffee can will keep swinging back and forth until someone stops it.



Your task is to think about out what happens to the pendulum **when you change the length of the string**.

Multiple choice:

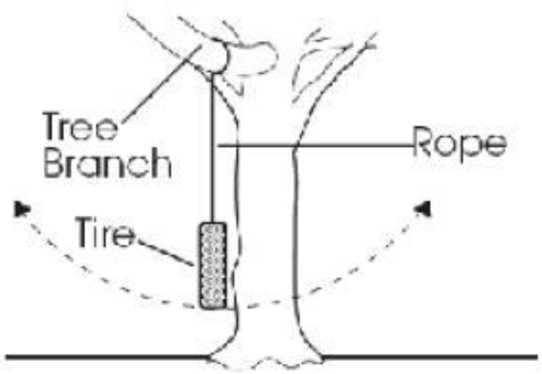
✓ OK

Ungraded open response:
B) Write a hypothesis you could test that states how the length of the string affects how far the pendulum goes.

Ungraded open response:
C) Describe how you would plan and conduct an experiment to test your hypothesis.

Tire Swing

A class investigating the motion of a tire swing collected the data in the table below. The students were able to draw conclusions about the factors that affect the motion of a swing. Two students from the class decide to use the class data to build a different-size tire swing in their backyard. They build the tire swing shown in the figure.



Tire Swing

Data table

Trial	Length of rope (meters)	Mass of tire (kilograms)	Time for tire to swing back & forth once (seconds)
1	2	10	2.8
2	2	20	2.8
3	4	10	4.0
4	4	20	4.0

After testing the swing, they decide that they want to make it swing faster. Based on the data from the class investigation, what could the students do to make their tire swing move back and forth faster?

Multiple choice:

- ✓ Use a shorter rope
- ✗ Use a longer rope
- ✗ Use a less massive tire
- ✗ Use a more massive tire

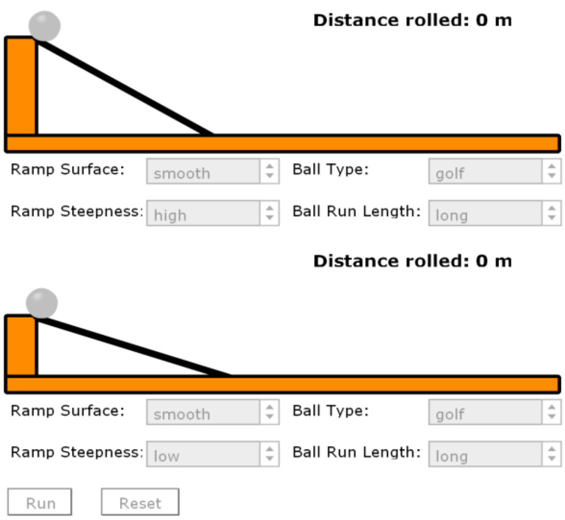
B) Explain your answer.

Ungraded open response:

Appendix B: Questions during Intervention

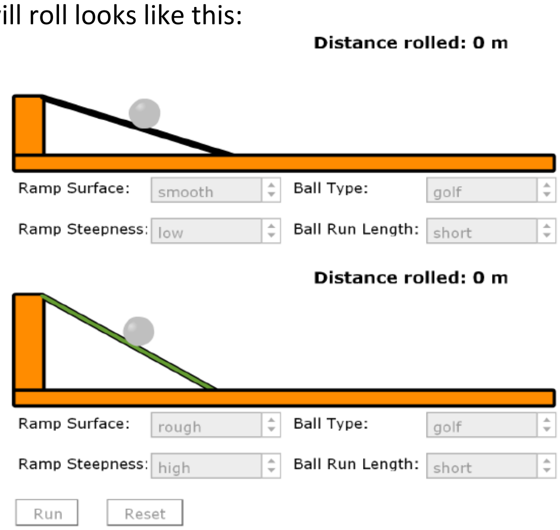
Direct-with-reify example

In the ramp experiment, you are testing if the variables (steepness, run length, the type of ball, and surface) affect how far the ball rolls. To figure out if one of these variables *by itself* affects how far the ball rolls, we use the **control of variables strategy**. The control of variables strategy says that we keep all variables the same and only change the one we want to test the effects of. This strategy lets us find how one variable alone does or does not affect the outcome of our experiment. For example, to test if **steepness** affects our outcome (how far the ball rolls) we should set up both ramps like this:



The only difference between the two ramps is the steepness. The run length, ball type, and surface are the same.

A setup that does not use the control of variables strategy to test if **steepness** affects how far the ball will roll looks like this:



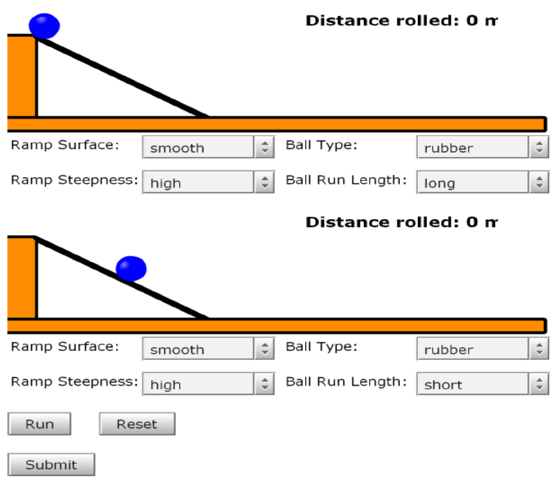
Even though the steepness is different, at least one of the other variable values is different between the two setups. In this case, the surface type is "smooth" in the upper setup and "rough" in the lower setup.

This means that this variable may also influence the results, not just steepness. Thus, if we were to run this experiment, we cannot tell for sure if steepness alone affects how far the ball rolls in this case.

Therefore, always remember that to test if only one factor affects an outcome, you should use the **control of variables strategy**. This means you should only change just that one variable and make sure all others are the same in the two setups.

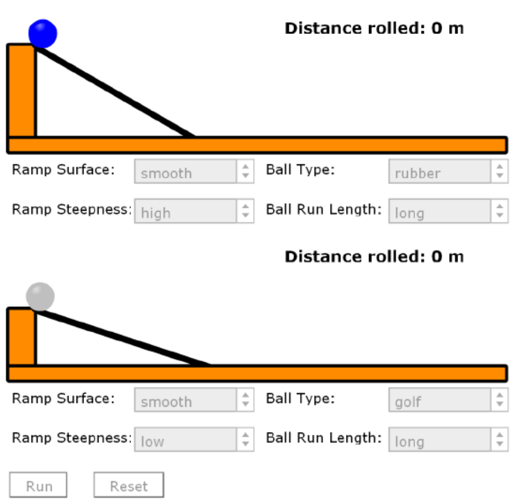
Discovery learning condition example

Set up the ramps to test if **run length** affects how far the ball rolls. You may change your setup as many times as you like and press "run" to test the setup.



Direct-without-reify

Look at the ramp setups below. Will these setups tell you for sure if steepness affects how far the ball will roll?



Will these setups tell you for sure if steepness affects how far the ball will roll?
Multiple choice:

☒ yes
☑ no

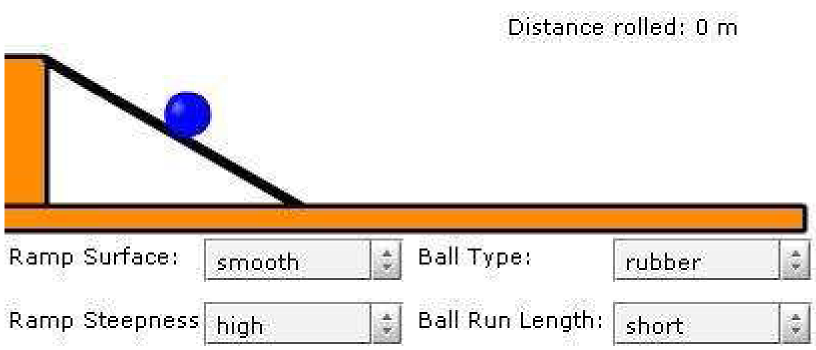
b) The experiment **cannot** tell you for sure if steepness determines how far the ball will roll. In order to do so, steepness needs to be the only variable that is different between the two ramps and all other variables should be the same for the two ramps.

Press RUN to see which other variables are different.

Appendix C: Ramp Post test

Based on the ramp experiments you just did, explain how you can determine if one particular variable affects how far the ball rolls.
Ungraded open response:

41) Again, keep the ramp experiment in mind.
When there are many variables that can be changed, explain how you can determine how **each variable** affects the distance the ball rolls.
Ungraded open response:



Recall that the ramp has four variables you can change:
- The steepness can be either a HIGH angle or a LOW angle.
- The run length can either be a SHORT run or a LONG run.
- The ball type can either be a RUBBER ball or a GOLF ball.
- The surface can either be a SMOOTH black surface or a ROUGH green surface.

Based on your experiments, what combination of variable settings would cause the ball to roll the **farthest**? Why?
Ungraded open response:

b) What combination of variable settings would cause the ball to roll the **shortest**? Why?
Ungraded open response:

Appendix D: Statistical Analysis

GET FILE='C:\Documents and Settings\falcor\My Documents\Science Assistments\Klahr Ramp Experiment Assistment\Fitchburg Results\day1 da y2\Patrick Analysis\fitchburgdata.sav'. DATASET ACTIVATE DataSet1. RECODE POST_Inquiry4_3 (0=1) (1=0) (2=0) (3=0) (4=0) INTO POST_Inquiry4_3recode. EXECUTE. RECODE POST_Inquiry5_3 (0=1) (1=0) (2=0) (3=0) (4=0) INTO POST_Inquiry5_3recode. EXECUTE. COMPUTE POST_Open_CVS=POST_Inquiry4_3recode+POST_Inquiry4_5+POST_Inquiry5_3recode+POST_Inquiry5_5 +POST_Inquiry6_4+POST_Inquiry6_5+PO ST_Inquiry7_4+POST_Inquiry7_5. EXECUTE. COMPUTE POST_Open_Content=POST_Inquiry6_2+POST_Inquiry7_2. EXECUTE. GLM POST_Open_CVS POST_Open_Content BY CONDITION /METHOD=SSTYPE(3) /INTERCEPT=INCLUDE /POSTHOC=CONDITION(TUKEY) /EMMEANS=TABLES(CONDITION) /PRINT=DESCRIPTIVE ETASQ OPOWER RSSCP HOMOGENEITY /CRITERIA=ALPHA(.05) /DESIGN= CONDITION.

COMPUTE total_inquiry_pre=PRE_Inquiry4_CVS+PRE_Inquiry6_CVS+PRE_Inquiry9_CVS. EXECUTE. COMPUTE total_ramp_pre=PRE_Ramp1_NC+PRE_Ramp2_MC+PRE_Ramp3_SC+PRE_Ramp4_MC. EXECUTE. GLM POST_Open_CVS POST_Open_Content BY CONDITION WITH total_inquiry_pre total_ramp_pre /METHOD=SSTYPE(3) /INTERCEPT=INCLUDE /EMMEANS=TABLES(CONDITION) WITH(total_inquiry_pre=MEAN total_ramp_pre=MEAN) COMPARE ADJ(BONFERRONI) /PRINT=DESCRIPTIVE ETASQ OPOWER RSSCP HOMOGENEITY /CRITERIA=ALPHA(.05) /DESIGN=total_inquiry_pre total_ramp_pre CONDITION.

General Linear Model

Notes

	Output Created	01-May-2009 13:06:55
	Comments	
Input	Data	C:\Documents and Settings\falcor\My Documents\Science Assistments\Klahr Ramp Experiment Assistment\Fitchburg Results\day1 day2\Patrick Analysis\fitchburgdata.sav
	Active Dataset	DataSet1
	Filter	filter_$ IEP=0 & ThrowAway=0 (FILTER)
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	133
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data for all variables in the model.

	Syntax	GLM POST_Open_CVS POST_Open_Content BY CONDITION WITH total_inquiry_pre total_ramp_pre /METHOD=SSTYPE(3) /INTERCEPT=INCLUDE /EMMEANS=TABLES(CONDITION) WITH(total_inquiry_pre=MEAN total_ramp_pre=MEAN) COMPARE ADJ(BONFERRONI) /PRINT=DESCRIPTIVE ETASQ OPOWER RSSCP HOMOGENEITY /CRITERIA=ALPHA(.05) /DESIGN=total_inquiry_pre total_ramp_pre CONDITION.
Resources	Processor Time	0:00:00.047
	Elapsed Time	0:00:00.046

[DataSet1] C:\Documents and Settings\falcor\My Documents\Science Assistments\Klahr Ramp Experiment Assistment\Fitchburg Results\day1 day2\Patrick Analysis\fitchburgdata.sav

Between-Subjects Factors

		Value Label	N
CONDITION	A	Direct Reify	37
	B	Discovery	29
	C	Direct No Reify	35

Descriptive Statistics

	CONDITION	Mean	Std. Deviation	N
POST_Open_CVS	A Direct Reify	4.4054	2.15329	37
	B Discovery	4.8276	2.42117	29
	C Direct No Reify	3.6286	1.76711	35
	Total	4.2574	2.14781	101
POST_Open_Content	A Direct Reify	4.5135	3.23713	37
	B Discovery	5.5517	3.01882	29
	C Direct No Reify	4.9143	3.35517	35
	Total	4.9505	3.21365	101

Box's Test of Equality of Covariance Matrices[a]

Box's M	5.563
F	.899
df1	6
df2	181125.810
Sig.	.494

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + total_inquiry_pre + total_ramp_pre + CONDITION

Bartlett's Test of Sphericity[a]

Likelihood Ratio	.000
Approx. Chi-Square	47.975
df	2
Sig.	.000

Tests the null hypothesis that the residual covariance matrix is proportional to an identity matrix.

a. Design: Intercept + total_inquiry_pre + total_ramp_pre + CONDITION

Multivariate Tests[d]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.483	44.289[a]	2.000	95.000	.000
	Wilks' Lambda	.517	44.289[a]	2.000	95.000	.000
	Hotelling's Trace	.932	44.289[a]	2.000	95.000	.000
	Roy's Largest Root	.932	44.289[a]	2.000	95.000	.000
total_inquiry_pre	Pillai's Trace	.018	.889[a]	2.000	95.000	.415
	Wilks' Lambda	.982	.889[a]	2.000	95.000	.415
	Hotelling's Trace	.019	.889[a]	2.000	95.000	.415
	Roy's Largest Root	.019	.889[a]	2.000	95.000	.415
total_ramp_pre	Pillai's Trace	.048	2.413[a]	2.000	95.000	.095
	Wilks' Lambda	.952	2.413[a]	2.000	95.000	.095
	Hotelling's Trace	.051	2.413[a]	2.000	95.000	.095
	Roy's Largest Root	.051	2.413[a]	2.000	95.000	.095
CONDITION	Pillai's Trace	.091	2.283	4.000	192.000	.062
	Wilks' Lambda	.911	2.266[a]	4.000	190.000	.064
	Hotelling's Trace	.096	2.248	4.000	188.000	.065
	Roy's Largest Root	.065	3.098[c]	2.000	96.000	.050

a. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Multivariate Tests[d]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.483	44.289[a]	2.000	95.000	.000
	Wilks' Lambda	.517	44.289[a]	2.000	95.000	.000
	Hotelling's Trace	.932	44.289[a]	2.000	95.000	.000
	Roy's Largest Root	.932	44.289[a]	2.000	95.000	.000
total_inquiry_pre	Pillai's Trace	.018	.889[a]	2.000	95.000	.415
	Wilks' Lambda	.982	.889[a]	2.000	95.000	.415
	Hotelling's Trace	.019	.889[a]	2.000	95.000	.415
	Roy's Largest Root	.019	.889[a]	2.000	95.000	.415
total_ramp_pre	Pillai's Trace	.048	2.413[a]	2.000	95.000	.095
	Wilks' Lambda	.952	2.413[a]	2.000	95.000	.095
	Hotelling's Trace	.051	2.413[a]	2.000	95.000	.095
	Roy's Largest Root	.051	2.413[a]	2.000	95.000	.095
CONDITION	Pillai's Trace	.091	2.283	4.000	192.000	.062
	Wilks' Lambda	.911	2.266[a]	4.000	190.000	.064
	Hotelling's Trace	.096	2.248	4.000	188.000	.065
	Roy's Largest Root	.065	3.098[c]	2.000	96.000	.050

a. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Design: Intercept + total_inquiry_pre + total_ramp_pre + CONDITION

Multivariate Tests[d]

Effect		Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Intercept	Pillai's Trace	.483	88.577	1.000
	Wilks' Lambda	.483	88.577	1.000
	Hotelling's Trace	.483	88.577	1.000
	Roy's Largest Root	.483	88.577	1.000
total_inquiry_pre	Pillai's Trace	.018	1.777	.199
	Wilks' Lambda	.018	1.777	.199
	Hotelling's Trace	.018	1.777	.199
	Roy's Largest Root	.018	1.777	.199
total_ramp_pre	Pillai's Trace	.048	4.826	.476
	Wilks' Lambda	.048	4.826	.476
	Hotelling's Trace	.048	4.826	.476
	Roy's Largest Root	.048	4.826	.476
CONDITION	Pillai's Trace	.045	9.132	.658
	Wilks' Lambda	.046	9.062	.654
	Hotelling's Trace	.046	8.992	.650
	Roy's Largest Root	.061	6.196	.584

b. Computed using alpha = .05

d. Design: Intercept + total_inquiry_pre + total_ramp_pre + CONDITION

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
POST_Open_CVS	1.598	2	98	.208
POST_Open_Content	1.524	2	98	.223

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + total_inquiry_pre + total_ramp_pre + CONDITION

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F
Corrected Model	POST_Open_CVS	30.324[a]	4	7.581	1.689
	POST_Open_Content	85.814[c]	4	21.454	2.175
Intercept	POST_Open_CVS	395.243	1	395.243	88.039
	POST_Open_Content	371.701	1	371.701	37.683
total_inquiry_pre	POST_Open_CVS	4.479	1	4.479	.998
	POST_Open_Content	16.513	1	16.513	1.674
total_ramp_pre	POST_Open_CVS	.975	1	.975	.217
	POST_Open_Content	41.983	1	41.983	4.256
CONDITION	POST_Open_CVS	22.912	2	11.456	2.552
	POST_Open_Content	29.536	2	14.768	1.497
Error	POST_Open_CVS	430.983	96	4.489	
	POST_Open_Content	946.938	96	9.864	
Total	POST_Open_CVS	2292.000	101		
	POST_Open_Content	3508.000	101		
Corrected Total	POST_Open_CVS	461.307	100		
	POST_Open_Content	1032.752	100		

a. R Squared = .066 (Adjusted R Squared = .027)

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F
Corrected Model	POST_Open_CVS	30.324[a]	4	7.581	1.689
	POST_Open_Content	85.814[c]	4	21.454	2.175
Intercept	POST_Open_CVS	395.243	1	395.243	88.039
	POST_Open_Content	371.701	1	371.701	37.683
total_inquiry_pre	POST_Open_CVS	4.479	1	4.479	.998
	POST_Open_Content	16.513	1	16.513	1.674
total_ramp_pre	POST_Open_CVS	.975	1	.975	.217
	POST_Open_Content	41.983	1	41.983	4.256
CONDITION	POST_Open_CVS	22.912	2	11.456	2.552
	POST_Open_Content	29.536	2	14.768	1.497
Error	POST_Open_CVS	430.983	96	4.489	
	POST_Open_Content	946.938	96	9.864	
Total	POST_Open_CVS	2292.000	101		
	POST_Open_Content	3508.000	101		
Corrected Total	POST_Open_CVS	461.307	100		
	POST_Open_Content	1032.752	100		

a. R Squared = .066 (Adjusted R Squared = .027)

c. R Squared = .083 (Adjusted R Squared = .045)

Tests of Between-Subjects Effects

Source	Dependent Variable	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	POST_Open_CVS	.159	.066	6.755	.501
	POST_Open_Content	.078	.083	8.700	.622
Intercept	POST_Open_CVS	.000	.478	88.039	1.000
	POST_Open_Content	.000	.282	37.683	1.000
total_inquiry_pre	POST_Open_CVS	.320	.010	.998	.167
	POST_Open_Content	.199	.017	1.674	.249
total_ramp_pre	POST_Open_CVS	.642	.002	.217	.075
	POST_Open_Content	.042	.042	4.256	.533
CONDITION	POST_Open_CVS	.083	.050	5.104	.499
	POST_Open_Content	.229	.030	2.994	.312

b. Computed using alpha = .05

Residual SSCP Matrix

		POST_Open_CVS	POST_Open_Content
Sum-of-Squares and Cross-Products	POST_Open_CVS	430.983	348.783
	POST_Open_Content	348.783	946.938
Covariance	POST_Open_CVS	4.489	3.633
	POST_Open_Content	3.633	9.864
Correlation	POST_Open_CVS	1.000	.546
	POST_Open_Content	.546	1.000

Based on Type III Sum of Squares

Estimated Marginal Means

CONDITION

Estimates

Dependent Variable	CONDITION	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
POST_Open_CVS	A Direct Reify	4.336[a]	.353	3.634	5.037
	B Discovery	4.866[a]	.395	4.082	5.650
	C Direct No Reify	3.671[a]	.360	2.956	4.386
POST_Open_Content	A Direct Reify	4.319[a]	.524	3.279	5.359
	B Discovery	5.685[a]	.585	4.523	6.847
	C Direct No Reify	5.009[a]	.534	3.949	6.069

a. Covariates appearing in the model are evaluated at the following values: total_inquiry_pre = 1.4950, total_ramp_pre = 1.3564.

Pairwise Comparisons

				a	
Dependent Variable	(I) CONDITION	(J) CONDITION	Mean Difference (I-J)	Std. Error	Sig.[a]
POST_Open_CVS	A Direct Reify	B Discovery	-.530	.533	.968
		C Direct No Reify	.665	.509	.584
	B Discovery	A Direct Reify	.530	.533	.968
		C Direct No Reify	1.195	.533	.082
	C Direct No Reify	A Direct Reify	-.665	.509	.584
		B Discovery	-1.195	.533	.082
POST_Open_Content	A Direct Reify	B Discovery	-1.366	.791	.262
		C Direct No Reify	-.690	.754	1.000
	B Discovery	A Direct Reify	1.366	.791	.262
		C Direct No Reify	.676	.790	1.000
	C Direct No Reify	A Direct Reify	.690	.754	1.000
		B Discovery	-.676	.790	1.000

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

Pairwise Comparisons

Dependent Variable	(I) CONDITION	(J) CONDITION	95% Confidence Interval for Difference[a]	
			Lower Bound	Upper Bound
POST_Open_CVS	A Direct Reify	B Discovery	-1.830	.770
		C Direct No Reify	-.575	1.905
	B Discovery	A Direct Reify	-.770	1.830
		C Direct No Reify	-.103	2.493
	C Direct No Reify	A Direct Reify	-1.905	.575
		B Discovery	-2.493	.103
POST_Open_Content	A Direct Reify	B Discovery	-3.293	.561
		C Direct No Reify	-2.528	1.148
	B Discovery	A Direct Reify	-.561	3.293
		C Direct No Reify	-1.248	2.600
	C Direct No Reify	A Direct Reify	-1.148	2.528
		B Discovery	-2.600	1.248

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.091	2.283	4.000	192.000	.062	.045
Wilks' lambda	.911	2.266[b]	4.000	190.000	.064	.046
Hotelling's trace	.096	2.248	4.000	188.000	.065	.046
Roy's largest root	.065	3.098[c]	2.000	96.000	.050	.061

Each F tests the multivariate effect of CONDITION. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Multivariate Tests

	Noncent. Parameter	Observed Power[a]
Pillai's trace	9.132	.658
Wilks' lambda	9.062	.654
Hotelling's trace	8.992	.650
Roy's largest root	6.196	.584

Each F tests the multivariate effect of CONDITION. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

Univariate Tests

Dependent Variable		Sum of Squares	df	Mean Square	F	Sig.
POST_Open_CVS	Contrast	22.912	2	11.456	2.552	.083
	Error	430.983	96	4.489		
POST_Open_Content	Contrast	29.536	2	14.768	1.497	.229
	Error	946.938	96	9.864		

The F tests the effect of CONDITION. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

Univariate Tests

Dependent Variable		Partial Eta Squared	Noncent. Parameter	Observed Power[a]
POST_Open_CVS	Contrast	.050	5.104	.499
POST_Open_Content	Contrast	.030	2.994	.312

The F tests the effect of CONDITION. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05