Assessing Students' Understanding of Controlling Variables

Arhonda Gogos and George E. DeBoer
AAAS Project 2061

National Association for Research in Science Teaching
Annual Conference

New Orleans, LA

April 15-18, 2007

Two studies are reported here that attempt to gain additional information about how middle school students think about controlling variables by using multiple-choice items in which common misconceptions are used as distractors. The results show that a number of students selected correct answers based on the knowledge they had of the context of the item rather than ideas they had about controlling variables, and that even students who seemed to have an understanding of controlling variables often did so unsystematically and often in combination with ideas about the context. The results show that multiple choice questions that incorporate misconceptions as distractors can be a useful way of uncovering what students know about controlling variables, especially when multiple items are used. Furthermore, initial inspection of student data from a second study suggests that multiple choice questions can be even more informative when combined with a follow-up multiple choice question that asks students to justify their answer selections.

Introduction

One important aspect of scientific inquiry involves deliberately controlling conditions when conducting scientific investigations. *Science for All Americans* says: "By varying just one condition at a time, [scientists] can hope to identify its exclusive effects on what happens, uncomplicated by changes in other conditions" (AAAS, p. 4). For students to be literate in science they should appreciate the logic of controlling variables and recognize what can be determined and what cannot be determined when appropriate controls are or are not in place. It is important to note that in *Science for All Americans*, the goal is for individuals to become knowledgeable about how science operates and to become critical consumers of scientific claims, not necessarily to be able to conduct scientific investigations themselves. Assessment can help educators probe ideas that students have about controlling variables (both correct and incorrect), and the results of assessment can be used to develop the instruction needed to focus student attention on the importance of "varying just one condition at a time."

One thing we know is that when faced with questions about controlling variables (CV), students often use ideas about the physical situation being described rather than drawing on the idea that only one variable should vary while all others are kept constant (for a review see Zimmerman 2005). In other words, students will decide (sometimes correctly and sometimes incorrectly) that a particular variable does or does not have an effect on an experimental outcome rather than that the variable needs to be *held constant* in order to find out the effect of another variable on the

outcome. The literature on CV also suggests that until their knowledge of CV is secure, students may use a combination of these valid and invalid strategies. This has made testing of student understanding of controlling variables difficult because it is possible that students may choose the correct answer to a multiple choice question on the basis of what they think will happen in a particular experiment, not on the basis of what they know about controlling variables.

Two studies are reported here that attempt to gain additional information about how middle school students think about controlling variables and to develop assessment strategies to reveal the ideas they have. The assessment items that we use are all multiple-choice items in which common misconceptions are used as distractors. This narrows the choices that students have to ideas that are either correct or that have been identified in the literature as problematic for students. The items are also closely aligned with learning goals that have been clearly articulated.

The first study confirms the well documented tendency for students to use what they think they know about a situation when answering this type of question. We observed this tendency not only in students who seem to have little if any understanding of CV, but also in students whose responses suggest that their knowledge of CV is relatively secure. In the second study, we added a follow-up multiple choice question to each question about controlling variables. In this follow-up question, students indicated why they selected the answers they did in the first question. In that study as well, we found that students often used their knowledge of the context of the experiment when drawing conclusions about what the experiment was designed to find out.

This work is part of a larger project to develop standards-based, distractor-driven, multiple-choice assessment items that target ideas in a number of different science content areas, as well as the nature of science. This particular aspect of the work has implications for how to test students' understanding of CV and how much confidence we can have in the results of testing in this area.

Procedures

Item Development. The procedure we use to develop assessment items involves three stages: (1) clarifying the targeted learning goal, (2) designing assessment tasks that are precisely aligned to the specific ideas in the targeted learning goal, and (3) revising items based on data obtained from interviewing and/or pilot-testing items with students. The process aims to achieve a precise link between a targeted idea and the test item itself.

Key ideas. Although state and national content standards provide important guidance to assessment developers regarding what students should know in science, these statements often do not provide enough precision regarding what students should know, which is needed to assess the students effectively. To increase the precision of the alignment between content and assessment, we further subdivide the content standards into finer-grained statements of knowledge, or *key ideas*, and then we clarify each key idea by indicating what it is that we expect students to know about that idea and what the boundaries of that knowledge are for purposes of assessment. In the case of controlling variables, the key idea says:

If more than one variable changes at the same time in an experiment, the outcome of the experiment may not be clearly attributable to any one of the variables (1B/M2a) (AAAS, 1993, p. 12).

Clarification statements. Key ideas are then elaborated to specify exactly what we expect students to know about the key idea. The following is the clarification statement for controlling variables:

Students should know that by varying more than one variable at a time it is not possible to determine the correlation between either variable and the outcome of an experiment. Students should also know that by changing a single variable at a time and holding all other relevant variables constant it is possible to determine whether that single variable is correlated with the outcome or not.

Students are not expected to know which variables, out of all possible variables, could be related to the outcome. Nor are students expected to know that it may not be possible to control or even identify all relevant variables. These ideas are included in 1B/M2b and 1B/H3. However, when given a set of variables, students are expected to know that in order to determine if there is a relationship between a particular variable and an outcome, all other variables in the set must remain constant.

Students should know that the reason for controlling a particular variable (holding it constant) in an experiment is because it may have an effect on what is being tested. Students are expected to apply this knowledge in contexts that involve individual events or objects and groups of people, events, or objects. Students are not expected to know when they can or cannot generalize beyond the given experimental and control groups.

Students should know that a variable is an entity that may assume different values, either quantitative or qualitative. Students should know that "variable" does not refer to a particular value of a variable. For example, they should know that when "types of liquid" is defined as a variable, water and juice are "values" of the variable and not variables themselves. Students are not expected to know the terms "independent" and "dependent" variables. Students should know what a controlled experiment is, and what a control group and an experimental group are in a controlled experiment.

Pilot Testing. Pilot testing is used to find out from students if anything is confusing about a test item, if there are words they do not understand, and why they think each answer choice is correct or incorrect. Students are told that they are helping to improve testing in science by providing complete and honest answers to questions about those items. In addition to being asked to explain why each answer choice is correct or incorrect, they are also asked to comment on the usefulness of graphs and pictures and to offer suggestions of other answer choices that could be used.

During pilot-testing, students respond in writing to the following questions about each item:

1. Is there anything about this test question that was confusing? Explain.

2. Circle any words on the test question you don't understand or aren't familiar with.
3. Is answer choice A correct? Explain why. Yes No Not Sure
4. Is answer choice B correct? Explain why. Yes No Not Sure
5. Is answer choice C correct? Explain why. Yes No Not Sure
6. Is answer choice D correct? Explain why. Yes No Not Sure
7. Did you guess when you answered the test question? Yes No
8. Please suggest additional answer choices that could be used.
9. Was the picture or graph helpful? If there was no picture or graph, would you like to see one?
10. Have you studied this topic in school? Yes No Not Sure
11. Have you learned about it somewhere else? Where? Yes No Not Sure

From their written responses, we are able to determine: (1) whether students used the targeted idea to answer the question or if they used other knowledge or test taking strategies instead, (2) whether the item is comprehensible to them, and (3) whether the answer choices are plausible to them. Mismatches between the answer choices that students select and the reasons they give provide information about whether the item is likely to yield false positive or false negative results. Student responses also provide information regarding student misconceptions.

The Test Items. For our work on controlling variables, we designed assessment items of four types, each of which probes students' understanding of CV in slightly different ways. The four types are:

I. Given an idea to be tested (hypothesis) and an experimental setup, explain *why* certain variables are (or should be) kept constant.
II. Choose an experimental setup to test a certain idea (hypothesis).
III. Choose an idea (hypothesis) that could be tested by using a particular controlled experimental setup.
IV. Given an experiment with two variables changing at the same time, determine that no conclusion can be drawn regarding the effect of each variable.

For each of the four types, we developed items in different contexts (physical science, life science, and consumer products). Contexts were chosen to address science ideas from an earlier (3-5) grade band so that it was reasonable to expect that the students would understand the science involved.

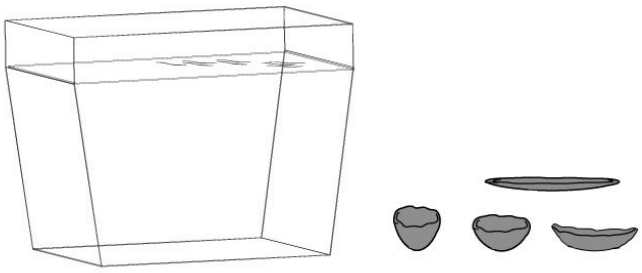
The studies reported here focus on data from type III items only. Most previous research on student understanding of CV has used items that ask students to choose an experimental setup to test a hypothesis (our type II). In the first study, students were given three questions that asked them to indicate what was being tested in an experiment. The items all used a common format, having one dependent variable and three independent variables. In a second study, half of the students were given one of the three test items and half were also given a follow-up multiple choice test item paired with the first item. The follow-up question asked why they had selected the answer choice that they did on the first item.

Study 1

In the first study, sixty-six eighth grade students in a suburban public school in the northeast United States responded to all three test items. The students were from a school district in which 10% of the students were eligible for free and reduced lunch. Students were heterogeneously placed in classrooms. Students received these questions along with others that we were testing as part of our general piloting of test items. The intent of pilot testing was to determine how successful students were in answering questions on each topic, if anything about the questions was confusing to students, and if the way the questions were written led to a significant number of false negative or false positive responses on the part of students. In what is reported here, we focus in particular on the possible false positive responses that result from students who answered the three questions above based on their knowledge of the context rather than on their understanding of controlling variables. All students were tested in May of 2006.

Item 1

A group of students uses clay to make boats of the same weight but different shapes. The students add 4 tablespoons of salt in a water tank and mix the water.



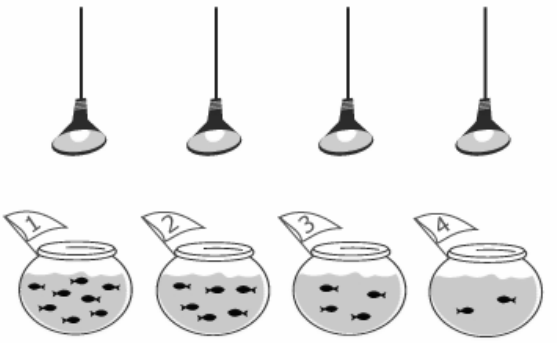
They place the boats in the tank and see if the boats float or if they sink.

What are the students testing with this experiment?

A. If shape affects the floating of the boats.
B. If weight affects the floating of the boats.
C. If shape and weight affect the floating of the boats.
D. If shape, weight, and the amount of salt in the water affect the floating of the boats.

Item 2

A student is interested in the behavior of fish. He has 4 fish bowls and 20 goldfish. He puts 8 fish in the first bowl, 6 fish in the second bowl, 4 fish in the third bowl and 2 fish in the fourth bowl. He places each fish bowl under light, he keeps the temperature at 25°C for all four bowls, and he observes the behavior of the fish.



Number of fish	8 fish	6 fish	4 fish	2 fish
Temperature	25°C	25°C	25°C	25°C

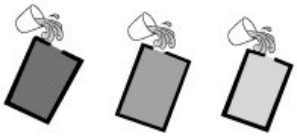
What is the student testing?

A. If the number of fish in the fish bowl affects the behavior of the fish.
B. If the temperature of the fish bowl affects the behavior of the fish.
C. If the temperature and the amount of light affect the behavior of the fish.
D. If the number of fish, the temperature, and the amount of light affect the behavior of the fish.

Item 3

A sailing team wants to select one out of three fabrics for their new uniform. They decide to do the following experiment:

They cut the same size pieces from each fabric and wet each piece with the same amount of water.



They hang the pieces in the sunlight and they check every 10 min to see if any of them is dry.

What is the team testing with this experiment?

A. If the amount of water affects how fast each fabric dries.
B. If the material of each fabric affects how fast the fabric dries.
C. If the material of each fabric and the amount of water affect how fast the fabric dries.
D. If the material of each fabric, the amount of water and the amount of light affect how fast the fabric dries.

Results:

The first question we addressed is how well do eighth grade students do on questions in which they have to evaluate an experimental setup to determine what is being tested in that experiment. Although each of the three test items was answered correctly by approximately two-thirds of the students (65.2%, 65.2%, and 69.7%), different combinations of students answered each of the items correctly. Only 45.5% of the students answered all three correctly, 19.7% answered two correctly, 24.2% answered one correctly, and 10.6% answered none correctly (see Table 1). This suggests that at least for some of the students with some understanding of CV, their knowledge of CV is not secure enough to answer correctly from one item to the next.

Table 1. Percent of students with 0, 1, 2, or 3 items correct.

Item	Percent Correct (n=66)
All three correct	45.5%
Two correct	19.7%
One correct	24.2%
Zero correct	10.6%

The second question we addressed is what the students' written comments say about why they selected the answer choices that they did. Based on the written comments students made about their answer choices, we were able to determine that a number of students who answered correctly used their knowledge of the context when answering the questions. This was true for all three items. Because these students' responses suggested that they were using knowledge of the context instead of their knowledge of controlling variables, we considered these to be possible false positive responses.

For the 30 students who selected the correct answer for *all three* items, four students (13.3%) used ideas about the specific context in some or all of their responses. For the 13 students who

answered correctly for *two* items, two students (15.4%) used ideas about the context. For the 16 students who answered correctly for *one* item, three students (18.8%) used ideas about the context. For the seven students who answered incorrectly for all three items, three students (42.9%) used ideas about the context (see Table 2).

Table 2. Using Knowledge of Context by Number of Items Correct

Number of Items Correct	Number of Students	Number Who Used Knowledge of Context	Percent Who Used Knowledge of Context
3	30	4	13.3%
2	13	2	15.4%
1	16	3	18.8%
0	7	3	42.9%

The third question we addressed is whether there is a pattern to how students use knowledge of context in answering these questions. From their written comments, we were able to determine not only which students used knowledge of the context in selecting answer choices, but we also observed that the extent to which they used their knowledge of context vs. knowledge of controlling variables varied from question to question. In Tables 3, 4, and 5 are examples of explanations provided by students who got the three questions correct. The first example is of a student who did not use knowledge of context for any of the three items, and the other two examples are of students who made partial use of their knowledge of the context for the items.

Student One used the targeted idea (CV) for all three items without using any ideas about the context:

Table 3. Explanations Provided by Student One

Item	Explanation provided by student
Clay Boats	"All of the boats are the same weight every time and the water and salt is the same every time, but the shape varies."
Fish Bowls	"He varies the number of fish in the bowl, the # of fish is the variable." and "The temperature and light are the controls."
Fabrics	"The fabric is the variable." "The water and light is not the variable."

Student Two chose the correct answer for all three items but used prior ideas for one of the variables in item 1 and evaluated the other variable based on a correct understanding of CV; for items 2 and 3 the student did not draw on ideas about the context:

Table 4. Explanations Provided by Student Two.

Item	Explanation provided by student
Clay Boats	"All the boats weigh the same." "The amount of salt will not affect the boats from floating."
Fish Bowls	"[# of fish] is the only variable in your experiment." "The light & temp. is the same."
Fabrics	"They had different fabrics." "The water & light was the same."

Student Three answered each of the three items correctly and used ideas about the context in all items, but for only some of the variables:

Table 5. Explanations Provided by Student Three.

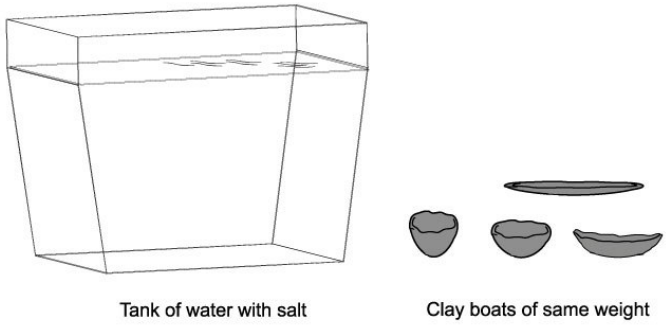
Item	Explanation provided by student
Clay Boats	"Shape can affect the boat." "Amount of salt shouldn't matter. All same weight."
Fish Bowls	"The number of fish in a fish bowl because they could eat each other, or reproduce." "No, heat/light don't affect the fish unless you change it."
Fabrics	"They all got the same amount of water poured on them." "The material of the fabric would affect how fast the water dries." "The amount of light has nothing to do with it."

Study 2

The purpose of the second study was to investigate additional reasons students may give for their answers to CV questions by using a follow-up multiple choice question. In this second study, one of the three items was randomly assigned to each of three groups of approximately 35 seventh grade students. Three other groups of 35 seventh grade students received one of the three items paired with a follow-up multiple choice test item that asked why they had selected the answer that they did to the first question. A total of 214 seventh grade students received one of three items either paired or unpaired with a follow-up question, for a total of about 35 students in each of six testing conditions. All items were randomly distributed within each classroom so that there were approximately four to five students from each of eight classrooms per testing condition. These seventh grade students were from the same school as the eighth grade students who took part in the first study. The items were pilot tested in February 2007. The items that seventh grade students received were slightly revised versions of what the eighth grade students received in study 1.

Item 1

A group of students uses clay to make boats of different shapes. All the boats they make have the same weight. The students add 4 tablespoons of salt in the water tank and mix the water. They place the boats in the tank and see if the boats float or if they sink.



What can the students find out from doing just this experiment?

A. If shape affects the floating of the boats.

B. If weight affects the floating of the boats.

C. If weight and amount of salt affect the floating of the boats.

D. If shape, weight, and the amount of salt in the water affect the floating of the boats.

Follow-up Item to Accompany Item 1

Why did you choose that answer?

E. Because I already know what affects the floating of the boats.

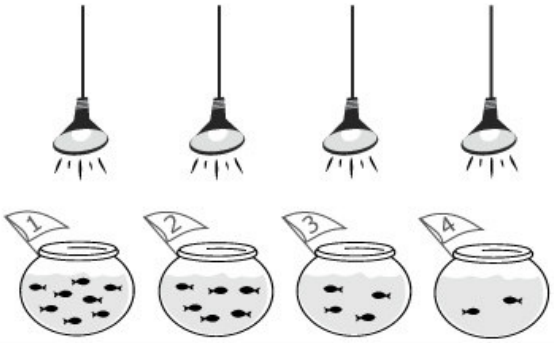
F. Because that is what is allowed to change in this experiment.

G. Because that is what stays the same in this experiment.

H. Because that is what the student decided to include in this experiment.

Item 2

A student is interested in the behavior of fish. He has 4 fish bowls and 20 goldfish. He puts 8 fish in the first bowl, 6 fish in the second bowl, 4 fish in the third bowl and 2 fish in the fourth bowl. He places each fish bowl under light, he keeps the temperature at 75°F for all four bowls, and he observes the behavior of the fish.



Number of fish	8 fish	6 fish	4 fish	2 fish
Temperature	75°F	75°F	75°F	75°F

What can the student find out from doing just this experiment?

A. If the number of fish in the fish bowl affects the behavior of the fish.

B. If the temperature of the fish bowl affects the behavior of the fish

C. If the temperature and the amount of light affect the behavior of the fish.

D. If the number of fish, the temperature, and the amount of light affect the behavior of the fish.

Follow-up Item to Accompany Item 2

Why did you choose that answer?

E. Because I already know what affects the behavior of fish.

F. Because that is what is allowed to change in this experiment.

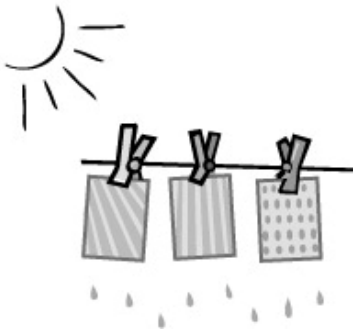
G. Because that is what stays the same in this experiment.

H. Because that is what the student decided to include in this experiment.

Item 3

A swimming team wants to select one of three fabrics for their new swimsuits. Each fabric is made of a different material. The team decides to do the following experiment:

They cut the same size pieces from each fabric and wet each piece with the same amount of water. They hang the pieces in the sunlight and they check every 2 minutes to see if any of the pieces are dry.



What can the team find out about the different fabrics from doing just this experiment?

A. If the amount of water affects how long it takes them to dry.

B. If the type of fabric affects how long it takes them to dry.

C. If the amount of water and the amount of light affect how long it takes them to dry.

D. If the type of fabric, the amount of water, and the amount of light affect how long it takes them to dry.

Follow-up Item to Accompany Item 3

Why did you choose that answer?

E. Because I already know what affects how long it takes them to dry.

F. Because that is what is allowed to change in this experiment.

G. Because that is what stays the same in this experiment.

H. Because that is what the student decided to include in this experiment.

Results:

As in Study 1, the first question we addressed was how well these students do on items in which they have to evaluate an experimental setup to determine what is being tested in that experiment. On all items and for all testing conditions combined, 46.3% (99/214) of the seventh grade students answered correctly (see Table 6). No significant differences were found for item type, or for item being paired or unpaired, and no significant item by pairing interaction effects were observed.

*Table 6. Percent Correct for Each Item (Study 2)

Item	Item Alone	Item when Paired with 2nd Level Item
Clay Boats	18/37 (48.7%)	23/37 (62.6%)
Fish Bowls	13/34 (38.2%)	13/34 (38.2%)
Fabrics	13/36 (36.1%)	19/36 (52.8%)
Total	44/107 (41.1%)	55/107 (51.4%)

*Note: The version of the paper on the Conference Proceedings CD has slightly different numbers, which have been corrected for this version of the paper. The changes do not change the statistical significance of the findings.

When the questions on controlling variables were paired with a follow-up question that asked students why they selected the answer choice that they did, 69.1% (38/55) of the students who selected the correct answer on the original item also selected the correct answer on the follow-up item (e.g., you can learn whether the shape of the clay boats matters because that was what was varied), 16.4% said they selected their answer based on the context (e.g., they already knew what affects the floating of the boats), 9.1% said they selected their answer on the basis of what was being controlled, and 5.5% said they selected their answer on the basis of what was being tested (see Table 7).

Table 7. Answer Selections on Follow-up Items for 55 Students who Selected Correct Answer on the Original Question.

Answer Choice	Clay Boats	Fish Bowls	Fabrics	Total	Percent
E. Already know what would happen	3	5	1	9	16.4%
F. This is what is allowed to change	17	9	12	38	69.1%
G. This is what stays the same	2	3	0	5	9.1%
H. This is what was included in the experiment	1	2	0	3	5.5%
Total	23	19	13	55	100%

*Note: The version of the paper on the Conference Proceedings CD has slightly different numbers, which have been corrected for this version of the paper. The changes do not change the statistical significance of the findings.

Seventh grade students were somewhat less likely to answer the type III items on controlling variables correctly than eighth grade students. Only 46.3% (99/214 from Table 6) of seventh grade students answered correctly, whereas the clay boat, fish bowl, and fabric items were answered correctly by 65.2%, 65.2%, and 69.7% of the eighth grade students, respectively. Seventh grade students based their correct answers on what they knew about the context of the question about as often as eighth grade students did—16.4% (9/55) for seventh grade students vs. 15.3% (9/59) for eighth grade students.

In Study 2, we also found that of the 15 students who indicated in their written comments on the original question that they had used knowledge of the context, nine of them (60%) also chose answer choice E on the follow-up question (I already knew what would happen). This suggests that a follow-up multiple choice question in which students are asked to give a reason for their

answer choice selection to a question about controlling variables detects at least some (60%) of the false positive responses that would otherwise not have been detected.

Discussion

Pilot testing of type III items revealed that each item was answered correctly by approximately two-thirds of the eighth grade students. However, different students answered each item correctly so that 45.5% of the students answered all three questions correctly and 10.6% of the students answered none of the three questions correctly. This means that 43.9% of the students got one or more questions correct and one or more incorrect, suggesting partial knowledge of CV. Therefore, although each of the items provides similar information about the group as a whole, each item taken individually can not be counted on to give an accurate picture of a student's understanding of the target idea. Student comments showed that a number of the students who selected the correct answers did so based on the knowledge they had of the context of the item. The results also indicate that students who used ideas they have about the context did so unsystematically and often in combination with ideas about controlling variables. This was observed even in students who answered correctly on all three items. When we added a follow-up multiple choice question to the questions about controlling variables to find out why students selected the answers they did, we found a 60% match between answer choice selections on the follow-up question and comments made on the original item. In other words, by using a follow-up multiple-choice question, we were able to uncover a significant number of false positive responses that would not be detected by using the original question alone.

These results show that multiple choice questions can provide useful information about what students know and do not know about controlling variables. Multiple choice testing is often criticized for testing only the facts of science, but multiple choice tests can be constructed to ask students to think through more complex situations and to explain and predict phenomena. Multiple choice questions can also be combined with follow-up questions that ask students to justify their answer selections and that are targeted to specific misconceptions that we are interested in testing. Although a considerable amount of effort is required to construct such test questions, when done well they provide educators with important information about what students know and can do. They also have the advantage of being able to focus student attention on particular aspects of the ideas that are being targeted and on the misconceptions that students are likely to hold. This is particularly useful for diagnosing gaps in student knowledge. Open-ended, free-response test items often produce answers that are unrelated to the idea being targeted, which makes it difficult to locate gaps in student understanding.

One other conclusion can be drawn from this research: When ideas about CV are developing, a single assessment item is unlikely to provide a stable measure of student understanding. It appears from this work that middle school is a transitional time for students in their understanding of controlling variables, sometimes drawing on ideas they have about CV and sometimes drawing on knowledge they have of the situation. Future research will study both older and younger students' understanding of controlling variables to determine how this understanding progresses over time.

This research is funded by the National Science Foundation Grant # ESI 0352473.

References

American Association for the Advancement of Science (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

American Association for the Advancement of Science (1990). *Science for All Americans*. New York: Oxford University Press.

Zimmerman, C. (2005). *The development of scientific reasoning: What psychologists contribute to an understanding of elementary science learning*. Paper commissioned by the National Academies of Science (National Research Council's Board of Science Education, Consensus Study on *Learning Science, Kindergarten through Eighth Grade*).