A Follow-Up Investigation of the Role of Cover Story on the Assessment of Experimental Design Skills

CSE Technical Report No. 594

Corinne Zimmerman
Learning Research and Development Center
University of Pittsburgh

Robert Glaser
CRESST / Learning Research and Development Center
University of Pittsburgh

April 2003

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Project 2.2. Classroom and Teachers' Assessment
Robert Glaser, Project Director, CRESST/Learning Research and Development Center, University of Pittsburgh

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, Office of Educational Research and Improvement, or the U.S. Department of Education.

A FOLLOW-UP INVESTIGATION OF THE ROLE OF COVER STORY ON THE ASSESSMENT OF EXPERIMENTAL DESIGN SKILLS

Corinne Zimmerman[1]
Learning Research and Development Center
University of Pittsburgh

Robert Glaser
CRESST/Learning Research and Development Center
University of Pittsburgh

Abstract

Cover story is a potentially relevant factor in the assessment of reasoning and problem solving in science, given repeated demonstration of its effect on laboratory tasks. This study follows up on a preliminary interview study that showed cover story influenced the way students were assessed at the end of an instructional unit. Two main changes characterize the present attempt to replicate earlier findings. First, testing materials were changed so as to reduce the number of possible explanations for the cover story effect, if found. Second, students completed the open-ended assessment in a group classroom setting rather than during individual interviews. There were no performance differences for cover stories that instructed students to design an experiment to test a *positive claim* (i.e., that coffee grinds are "good" for plants), a *negative claim* (i.e., that coffee grinds are "bad" for plants), or a *neutral* control. Observed differences, however, were related to the teacher that students had for the instructional unit. Implications for assessment are discussed.

Developing assessments to measure reasoning and problem solving continues to be a challenge for science educators. It has been suggested that the design of assessment tasks should be informed by cognitive theory and research (e.g., Baxter & Glaser, 1998). A pervasive and robust finding in the cognitive psychology literature is that the context within which a problem is presented can influence the way a problem is solved. For example, problems are often easier to solve when they are presented within a concrete or semantically rich cover story than in an abstract one (for a review, see Evans & Over, 1996). Given that cover story influences performance on laboratory tasks of reasoning and problem solving, we conducted a

[1] Corinne Zimmerman is now at the Department of Psychology, Illinois State University, Normal, IL 61790-4620.

preliminary study to ascertain whether cover story was a relevant factor for the assessment of reasoning and problem solving in the science classroom (Zimmerman, Glaser, & Raghavan, 2001).

In an initial study to investigate the role of cover story on the assessment of experimental design skills, we conducted one-on-one interviews with students in which they were asked to "think aloud" while completing an assessment task (Zimmerman & Glaser, 2001). Students' performance was assessed with an open-ended task administered upon completion of the *Experiments With Plants* curriculum unit (National Science Resource Center, 1992). The goal of this curriculum unit is "to teach students how to design and conduct controlled investigative experiments" (p. 1).

There were two versions of an end-of-unit assessment that required students to design an experiment to test a claim. One version included a cover story in which the claim was "positive" (i.e., test the claim that "coffee grinds are good for plants"), and the other version included a cover story in which the claim was "negative" (i.e., test the claim that "tap water is bad for plants"). This manipulation was selected because previous studies have shown that students perform differently in the laboratory when cover story materials have positive or negative outcomes for both control-of-variables tasks (Tschirgi, 1980) and for counterfactual reasoning tasks (e.g., German, 1999).

Overall, students in the negative claim condition outperformed those in the positive claim condition. They were more likely to design *controlled* experiments, and to suggest experiments in which they manipulated the variable of interest (i.e., type of water). Students in the positive claim condition were more likely to design *uncontrolled* experiments in which they did not chose the focal variable (i.e., coffee grinds) as the independent variable. That is, students in the two groups seemed to use different strategies for designing their experiments. Instead of testing the claim that coffee grinds are good for plants by setting up an experiment with green bean seeds (as suggested in both cover stories), students in the positive claim group acted as though they were *accepting* the claim and that their goal was to test the *generality* of it by using coffee grounds with a variety of plants (e.g., suggesting a design with green beans *vs.* sunflowers *vs.* pine trees).

Proposing designs with *plant type* (rather than coffee grinds) as the main independent variable is suggestive of the idea that students were attempting to

reproduce the "good effect" of coffee grinds but with different types of plants. This result is consistent with other studies showing that individuals treat "positive" and "negative" events differently (e.g., German, 1999; Roese, 1997; Tschirgi, 1980) and the interpretation that this tendency might reflect individuals' experience with everyday problem-solving situations in which the goal is to reproduce positive effects and eliminate negative effects (Tshirgi, 1980).

Although the pilot study showed that the cover story of an open-ended task could have an impact on how students are assessed (i.e., different versions may underestimate or overestimate students' abilities), it was necessary to complete a follow-up study for several reasons. First, if we are to make the claim that superficial changes in the wording of a task influence student performance, it is necessary to replicate the effect. Moreover, it is necessary to determine whether such a cover story effect would occur under normal classroom conditions. That is, the assessment of students by an individual teacher or by large-scale standardized tests (see, e.g., National Assessment of Educational Progress, 2001) typically is not done through one-on-one interviews. Rather, assessments are more commonly administered in a group context, with each student working individually. In order to generalize to the classroom situation, the cover story effect must be found outside of the laboratory.

Second, the cover story effect found in the preliminary study was open to a number of interpretations concerning *why* the effect might have occurred. It is possible that the valence of the claim influenced the way students approached the task (e.g., to reproduce positive effects and eliminate negative effects). The cover story materials were constructed such that the claims were worded as either positive or negative (i.e., "good" or "bad" for plants), but there are other plausible explanations to account for the performance difference. For example, the *familiarity* of the focal variables in the two cover stories could explain the performance differences. Water may be a more familiar variable than coffee grinds. Students' understanding of the role of water in a plants context may make this an inherently more familiar variable than coffee grinds and therefore is a plausible explanation that needs to be considered. It is necessary to tease apart the exact nature of the nuances in the cover story that could potentially lead to an over- or under-assessment of student performance.

In summary, cover story manipulations have been shown to influence students' performance on an open-ended experimentation task in a preliminary interview study (Zimmerman & Glaser, 2001). Although this study highlights the importance

of considering the effects of task variations such as cover story when designing instructional and assessment tasks, it is important to provide additional demonstration that such effects hold up in the classroom situation.

Two objectives guided the design of the present study. Our first objective was to determine whether or not we could replicate the cover story effect found in the interview study, and whether it would generalize to a group-administered, classroom-based assessment situation. Second, we wanted to ascertain the nature of the effect, if found. That is, we wanted to find out whether the valence of the claim or the familiarity of the materials was responsible for the performance differences found in the preliminary interview study.

Method

Students in six classrooms (N = 135) from a culturally diverse school in an urban center participated in the study (75 males, 60 females). Class sizes ranged from 20 to 24 students, with approximately equal gender ratios. Three teachers taught two science classes each. Teacher A (n = 48 students) was an experienced science teacher who had taught this particular curriculum unit for 5 years. Students from Teacher A's class participated in the preliminary study (Zimmerman & Glaser, 2001). Teacher B (n = 42 students) was an experienced science teacher who was teaching this unit for the first time in 3 years. Teacher C (n = 45 students) was an experienced teacher but inexperienced as a science teacher, and thus was teaching this unit for the first time.

A 2-page assessment was group administered to students. The task took approximately one class period to complete and required the students to design an experiment with plants. This task was intended to represent a typical end-of-unit assessment for the *Experiments With Plants* curriculum unit and was based on a task used previously in this school district (Raghavan, 1999). Three open-ended questions were asked: (a) Describe and explain how you would set up the experiment; (b) describe what you would measure; and (c) design a table to record the data collected throughout your experiment.

Three different versions of the plants task were created (see Appendix). Students were randomly assigned to one of these conditions. Each version had coffee grinds as the variable to be tested, to control for familiarity. In the *Positive Claim* version (n = 45), students were asked to design an experiment to find out whether coffee grinds are "good" for green bean plants, based on a character's

assertion that they are good for plants. In the *Negative Claim* version (n = 42), students were asked to design an experiment to find out whether coffee grinds are "bad" for green bean plants, based on a character's assertion. In the *Neutral* version (n = 45), students were asked to design an experiment to find out whether coffee grinds are "good or bad" for plants. This version did not include a character asserting a claim about coffee grinds (see Appendix).

A scoring scheme was devised to assess the specific skills considered to be important when learning about the scientific process (e.g., National Science Resource Center, 1992). A *process score* was calculated that included the following five components: (a) manipulating only one variable, (b) manipulating the correct variable, (c) keeping conditions constant (i.e., controls such as the amount of sunlight, soil, water, or coffee grinds), (d) use of repeated measurements (e.g., use of trials, averages, or multiple plants per condition), and (e) systematic observation (i.e., measurements made on a regular basis over time). Students received 0-2 points for each process skill and an overall process score (maximum of 10 points). This measure has been used in previous research to evaluate the science curriculum in this school district (Raghavan, 1999) and in the pilot project (Zimmerman & Glaser, 2001).

Results

A 3 (cover story) x 3 (teacher) analysis of variance (ANOVA) was used to analyze the process score. A main effect of cover story was not evident, $F(2, 123) = 1.04$, $p = 0.36$. The average process scores (maximum of 10 points) for the positive, neutral, and negative cover stories were 5.2, 5.4, and 4.8, respectively. There was, however, a main effect of teacher, $F(2, 123) = 14.25$, $p < .001$, with the students taught by Teachers A, B, and C scoring 6.7, 4.9, and 3.8, respectively (see Figure 1). The interaction between teacher and cover story was not significant, $F(4, 123) = 0.81$, $p = .52$.

Recall that the process score is comprised of five different skills. We examined each of these skills individually to see whether there were patterns not identifiable by the total process score. The percentage of students in each cover story condition receiving credit for these five skills appears in Table 1. Chi-square analyses were used to determine whether there was a relationship between each process skill (credit, no credit) and cover story condition (positive, negative, neutral), but in each case, these two variables were independent (χ^2s = 0.40 to 1.44; ps = 0.34 to 0.82).

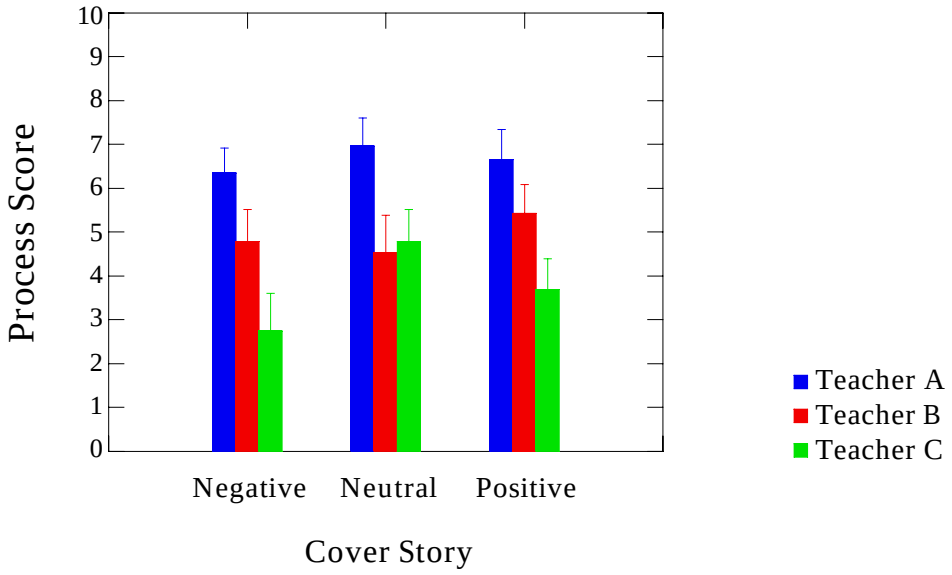

Figure 1. Mean process score by teacher and cover story.

Table 1

Percentage of Students in Each Condition Receiving Credit for Five Process Skills

Process skill	Positive	Negative	Neutral
Manipulate the correct variable	71.1	66.7	73.3
Manipulate only one variable	73.3	59.5	71.1
Use of repeated measures	46.7	50.0	53.3
Systematic observation	60.0	52.4	57.8
Keep conditions constant (controls)	46.7	59.5	53.3

Note. All comparisons were nonsignificant.

As can be seen in Table 2, there was a relationship between Teacher and receiving credit for four of the process skills: (a) manipulating the correct variable, χ^2 (2 *df*) = 6.46, $p < 0.05$; (b) manipulating only one variable, χ^2 (2 *df*) = 8.37, $p < 0.05$; (c) use of repeated measures, χ^2 (2 *df*) = 58.72, $p < 0.001$; and (d) keep conditions constant, χ^2 (2 *df*) = 9.73, $p < 0.01$. There was no relationship between Teacher and use of systematic observation, χ^2 (2 *df*) = 0.11, $p < 0.95$.

Table 2

Percentage of Students in Each Class Receiving Credit for Five Process Skills

Process skill	Teacher A	Teacher B	Teacher C
Manipulate the correct variable*	83.3	66.7	60.0
Manipulate only one variable*	81.3	69.3	53.3
Use of repeated measures***	93.8	30.8	20.0
Systematic observation	56.3	58.9	55.6
Keep conditions constant (controls)**	52.1	71.8	37.8

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Each student's design was characterized based on the main variable that was manipulated. The designs used by students taught by the three teachers are shown in Table 3. The type of experimental design was clearly related to teacher more than to the condition to which the student was randomly assigned. Statistical analyses are not possible due to low expected frequencies, but patterns are evident. As can be seen in Table 3, 20 of the students in Teacher A's classes (40%) proposed designs in which the *amount of coffee* was manipulated. Simple *coffee versus no coffee* designs were common for all students, but those in Teacher A's classes were more likely to use this design in conjunction with multiple trials. The reverse is seen for Teachers B and C, whose students were more likely to design experiments using a single plant for each condition.

Table 3

Numbers of Students Proposing Different Experimental Designs

Design	Teacher A (n = 48)	Teacher B (n = 39)	Teacher C (n = 45)
Coffee versus No coffee (multiple trials)	9	5	5
Coffee versus No coffee (single plant each)	2	13	18
Amount of coffee (with "No coffee" control)	10	2	—
Amount of coffee (without control)	10	1	—
Type of coffee (with "No coffee" control)	2	2	—
Coffee on all plants (multiple trials)	3	5	—
Other	4	3	2
No design	8	8	20

Note. "Other" included Type of compost, Amount of time the plant gets coffee, Number of fertilizer pellets, Coffee (present / absent) x Sunlight (present / absent), Number of seeds, Amount of space, Coffee vs. fertilizer, and Light vs. dark conditions.

The other striking difference was the number of students in Teacher C's classes who wrote descriptions that could not be classified as any design. The responses for these students typically were either procedures for planting a seed and taking care of the plant, or a procedure that involved putting coffee on a single plant (i.e., this does not meet the minimum requirements for an experiment generally or as taught in the curriculum unit).

Discussion

Our first objective was to determine whether cover story effects would occur in a group of students who were administered an end-of-unit assessment in the classroom rather than in the laboratory. Many laboratory studies, including our preliminary investigation (Zimmerman & Glaser, 2001), have shown that task variations can influence performance. In the real-world classroom, however, these effects get washed out by the situational aspects of a more authentic setting. Our second objective was to rule out alternate explanations for why the cover story effect occurred, if found.

There are a number of possible explanations for the absence of a cover story effect in the present study. First, it is possible that the revised versions of the cover story were not different enough. In the effort to have tightly controlled testing materials (i.e., control for familiarity, length, etc.), the manipulation may have been too subtle, such that the three cover stories differed by only a few words (see Appendix). Second, the previous findings in which a cover story effect was found could have been related to the *familiarity* of the variables rather than the valence of the claim (i.e., positive, negative). In the present study, familiarity was controlled for, but it may have been the factor driving the effect in the preliminary study. Third, the difference in cover story could be a very narrow influence when surrounded by other factors in the wider context of the class environment. We were interested to see whether the effect would generalize not just from one situation to another, but from the circumscribed environment of the laboratory to the environment of the real-world classroom. Clearly, factors beyond those of the testing materials were at work in the present study.

The most striking factor that influenced student performance in the classroom setting was the teacher who taught the curriculum unit. "Teacher effects" were not expected, so a formal documentation of teaching practice was not done. As noted previously, the teachers varied with respect to the amount of experience they had

with this particular curriculum unit (from 1 to 5 years). No interaction was found between teacher's experience and the cover story effect, so it is not the case that the effect is found for teachers with, for example, less teaching experience.

In general, the students in each classroom were learning different things from different teachers using the same standardized educational materials. Students in Teacher A's class learned about a range of experimental designs, the use of controls, and the importance of repeated measures when experimenting with plants. The students in Teacher C's class appear to have learned more about the procedures for planting seeds rather than using plants as a medium to learn about how to conduct investigations. The students in Teacher B's class were mixed with respect to whether the main skills learned were related to conducting investigations or keeping plants healthy.

The objective of the present study was to determine whether cover story variations that influenced student performance in the laboratory would generalize to a more authentic assessment situation. Although the main manipulation (cover story) did not result in differential performance by students, it is clear that it is important to consider the situational aspects of the classroom when conducting studies to determine whether local effects generalize from the laboratory to the richness of the classroom. In authentic situations, factors such as (a) teacher experience and performance, (b) students' knowledge, and (c) the performance environment may override and wash out the subtle manipulations that might exert an influence in tightly controlled settings.

References

Baxter, G. P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Issues and Practice, 17*, 37-45.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning.* Hove, UK: Psychology Press.

German, T. P. (1999). Children's causal reasoning: Counterfactual thinking occurs for "negative" outcomes only. *Developmental Science, 2*, 442-447.

National Assessment of Educational Progress. (2001). *The nation's report card: Science highlights 2000*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

National Science Resource Center. (1992). *Experiments with plants: Teacher's guide.* Burlington, NC: Carolina Biological Supply.

Raghavan, K. V. (1999). *Local systemic change: 1998-99 evaluation report for ASSET Inc.* Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133-148.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51,* 1-10.

Zimmerman, C., & Glaser, R. (2001). *Testing positive versus negative claims: A preliminary investigation of the role of cover story in the assessment of experimental design skills* (CSE Tech. Rep. No. 554). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Zimmerman, C., Glaser, R., & Raghavan, K. V. (2001, April). *The role of cover story in the assessment of experimental design skills.* Poster presented at the annual meeting of the American Educational Research Association, Seattle, WA.

APPENDIX

Cover Stories for the Plants Task

Positive Claim Cover Story

A fourth grader named Sonia was having breakfast one morning when she accidentally knocked the basket of used coffee grinds into the pot of a house plant. Her grandmother turned around and said, "Oh that's okay—coffee grinds are good for plants." This gave Sonia an idea. She wanted to do some experiments to find out if coffee grinds really are "good" for plants. She collected some green bean seeds for her experiment. Sonia asked you to help her set up the experiment.

Negative Claim Cover Story

A fourth grader named Sonia was having breakfast one morning when she accidentally knocked the basket of used coffee grinds into the pot of a house plant. Her grandmother turned around and said, "We better clean this up—coffee grinds are bad for plants." This gave Sonia an idea. She wanted to do some experiments to find out if coffee grinds really are "bad" for plants. She collected some green bean seeds for her experiment. Sonia asked you to help her set up the experiment.

Neutral Cover Story

A fourth grader named Sonia was having breakfast one morning when she accidentally knocked the basket of used coffee grinds into the pot of a house plant. This gave Sonia an idea. She wanted to do some experiments to find out if coffee grinds are good or bad for plants. She collected some green bean seeds for her experiment. Sonia asked you to help her set up the experiment.