JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 15, NO. 1, PP. 91-94 (1978)

COMMENTS AND CRITICISM

COMMENTS ON "THE EFFECT OF EXPERIENTIAL SCIENCE ON DEVELOPMENT OF LOGICAL THINKING IN CHILDREN"

RONALD G. GOOD, LEHMAN BARNES, TANNER GIRARD, PHILLIP HORTON, and GUY MORIN

Florida State University, Tallahassee, Florida 32306

This paper by Linn and Thier was selected by the JRST Award Committee as the outstanding article appearing in JRST for 1975. As authors of this critique, we assume that criteria for defining "outstanding" would include quality of research design as well as the more subjective area of importance or significance to the field of science education. No quarrel is taken with the latter criterion as applied to Linn and Thier's paper, but the research design, at least as reported in their paper, raises a number of issues and leaves many questions unanswered. Among the more important "design" questions are the following:

1. Why was the particular test (cart rolling down an inclined plane) chosen to assess logical thinking? About 17% of the energy sources unit (in terms of pages in the Teacher's Guide) is devoted to activities in which students roll objects down inclined planes. Since the experimental group has this equipment available to them while the controls did not, bias is introduced in terms of recognition of and familiarity with equipment shown in the test film and test sheet. Also, the question of just what is being measured, retention or developed ability, is raised since the experimental subjects were working with the inclined planes at the time of testing. This could have been controlled by scheduling the test five or six months after completion of the unit.

2. Why was a regression analysis not done to determine the variance in test scores that could be accounted for by the "treatment"? Such an analysis was done in an earlier SCIS study (Allen, 1973), and the variance accounted for by SCIS was 5% for cognitive scores and 3% for motivational scores, both very low considering the nature of the study.

3. What was the "control"? Certain assumptions might possibly be made regarding the general nature of the experimental treatment (SCIS), but no substantive information is available regarding the control classes. Since little is known about the SCIS classroom conditions and almost nothing is known about the control classroom conditions, even if all other aspects of the study had been controlled, what can be said about the possible "causal" factors? Could it be certain patterns of teacher behavior; the nature of the manipulative materials; the fact that students could manipulate materials; the nature of the interaction among the students; etc.?

4. Who were the "experimenters" who did the actual testing in the classrooms? If the teachers also did the testing, could it be possible that SCIS teachers, being anxious to show that their students were doing well, inadvertantly guided their students to more logical choices and reasons than did their control counterparts?

5. Why were all students' responses not scored "blindly"? The investigators, one of

Published by John Wiley & Sons, Inc. 0022-4308/78/0015-0091$01.00

whom played a significant role in the development of the SCIS program, apparently scored the students' test responses while knowing that each test represented a "control" or a "treatment." Having one member of the "staff" (SCIS?) score Part C of the tests hardly constitutes an effective control for the lack of blind scoring.

6. Do not the title and certain statements in the Summary and Conclusions overgeneralize the results (regardless of the "goodness" of the design)? For instance: "It was found that fifth-graders who studied Energy Scources appeared to be better logical thinkers than comparable fifth-grade students who did not. Also, students of *Energy Sources* had scores that were close to those of eighth-graders in comparable areas" (p. 60).

If we assume that each student spent about 25 hours working on the Energy Sources unit (two or three hours per week for 10 weeks), this means that the "average" student in the experimental group worked for about one school week on something designed to facilitate logical thought. Is it reasonable to assume that the equivalent of one week of such activity is sufficient to advance logical thinking from a fifth-grade level to an eighth-grade level?

7. How do the figures in Table I correspond to the figures given on page 57 under the heading, "Urban Fringe and Suburban Populations"? For example, how does a control mean of 0.37 (p. 57) yield a percentage score of 33 (Table I)? What specific numerical scores were given to the three categories of scoring described on pages 54 and 55?

We hope that Linn and Thier do not interpret these questions to mean that their study is unimportant. On the contrary, studies of this general nature are greatly needed to provide objective evidence that experiential science is superior to nonexperiential science, a belief that most science educators seem to hold. It seems critical, however, that such research studies follow reasonably rigorous design principles, especially for a study identified by researchers in science education as "most outstanding."

References

Allen, L. R. An evaluation of children's performance on certain cognitive, affective, and motivational aspects of the systems and subsystems unit of the Science Curriculum Improvement Study elementary science program. *Journal of Research in Science Teaching,* 1973, **10,** 125-134.

Linn, M. C., & Thier, H. D. The effect of experiential science on development of logical thinking in children. *Journal of Research in Science Teaching,* 1975, **12,** 49-62.

AUTHORS' RESPONSE

The critics of our paper admit that the research is "important" and "significant;" they dislike our research design. Of their seven criticisms, we disagree with three (numbers 2, 5, and 6), and three are points of information (numbers 3, 4, and 7). For criticism 1, our rationale given in the paper is not refuted.

1. Our rationale for choosing the cart experiment is given in the paper (Linn & Thier, 1975, p. 50, lines 8-24; p. 51, lines 12-18). In particular, we note "compensating variables ... is not explicitly taught in the unit" (Linn & Thier, 1975, p. 50), so retention is not an issue. The critics do not refute this. We disagree with them when they claim that the equipment for the cart experiment was familiar to experimental subjects; it was novel to all subjects. The similarity is that both experiments involve the height variable, a variable which over 95% of both experimental and control subjects could recognize and describe (Linn & Thier, 1975, p. 54).

2. We disagree with the critics in recommending a regression analysis for our data since the scores are dichotomous and not normally distributed. Rather, we question Allen's (1973) use of regression analysis for his largely categorical data. We feel that the magnitude of the effect on experimental classes can be easily grasped by inspection of our Table II. For example, the median score of control classes is 30% as compared to 45% for experimentals, a difference of 23% of the range.

3. Rationale for class selection is given in the paper (Linn & Thier, 1975, p. 56). The study did not assess specific factors contributing to the effect of *Energy Sources* (SCIS, 1971). Such studies would be of interest.

4. A point of information: either author Thier or the science coordinator served as the experimenter.

5. On this point, we disagree with the critics. The papers were scored blindly and the reliability check was done, as noted in the Acknowledgement (Linn & Thier, 1975, p. 62), by Warren Wollman who had just returned from three years of working in Geneva with Piaget.

6. We disagree. By scoring significantly higher on the cart experiment, the *Energy Sources* students did appear to be better logical thinkers than comparable controls, and scores were close for fifth-grade students of *Energy Sources* and eighth-graders in comparable areas.,

The critics' opinion of the "reasonableness" of our results is not supported by other research. We agree that the results are dramatic; that is why we reported them. They are not unreasonable and, in fact, parallel the findings of many researchers including those concerned with performance as opposed to competence (e.g., Olton & Crutchfield, 1969) and those involved in training (see for example, Case & Fry, 1973; Lawson & Wollman, 1976; Siegler, 1976).

7. This is a point of information. The means reported on page 57 are computed by giving equal weight to the eight areas involved in the study so that results can be generalized across areas. The means in Table I give equal weight to each class in the study but unequal weights to areas since different numbers of classes were gathered in each area. Of course, analysis by class would have yielded a higher t but its generalizability would have been limited.

Our critics find this paper "important" and "significant" but the design not "reasonably rigorous." The rationale for their criticism is not stated, but they appear to be uncomfortable with the nontraditional aspects of our design. We wonder whether they have considered the implications of their position. Many educators, psychologists, and philosophers are currently seeking alternatives to traditional designs which, after all, were borrowed from the biological sciences (see, for example, Bradley, 1968; Cronbach, 1975; Levine, 1975). We feel that the nontraditional design we used is an example of how traditional curriculum comparison studies described by Cronbach (1963) as too costly can be made cost-effective. If we had used a traditional design and randomly assigned teachers to treatments, it would have cost a minimum of $1000.00 per class for SCIS materials, inservice, and testing–the study involved 92 classes.

References

Allen, L. R. An evaluation of children's performance on certain cognitive, affective, and motivational aspects of the systems and subsystems unit of the Science Curriculum Improvement Study elementary science program. *Journal of Research in Science Teaching*, 1973, **10**, 125-134.

Bradley, J. V. *Distribution free statistical tests.* Englewood Cliffs, N.J.: Prentice Hall, 1968.

Case, R., & Fry, C. Evaluation of an attempt to teach scientific inquiry and criticism in a working class high school. *Journal of Research in Science Teaching,* 1973, **10**, 135-142.

Cronbach, L. J. Evaluation for course improvement. *Teachers College Record,* 1963, **64**, 672-683.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. *American Psychologist,* 1975, **30**, 116-127.

Lawson, A. E., & Wollman, W. T. Encouraging the transition from concrete to formal cognitive functioning—An experiment. *Journal of Research in Science Teaching,* 1976, **13**, 413-430.

Levine, M. Scientific method and the adversary model: Some preliminary thoughts. *American Psychologist,* 1974, **29**, 661-667.

Linn, M. C., & Thier, H. D. The effect of experiential science on development of logical thinking in children. *Journal of Research in Science Teaching,* 1975, **12**, 49-62.

Olton, R. M., & Crutchfield, R. S. Developing the skills of productive thinking. In Mussen, P., Langer, J., & Covington, M. V. (Eds.), *Trends and issues in developmental psychology.* New York: Holt, Rinehart & Winston, 1969, pp. 68-91.

Science Curriculum Improvement Study. *Energy sources. Teachers guide.* Chicago: Rand McNalley & Co., 1971.

Siegler, R. S. Three aspects of cognitive development. *Cognitive Psychology,* 1976, **8**, 481-520.