

In-Service Institutes: Do They Make a Difference?¹

ROBERT L. UFFELMAN,* A. JON MAGOON
PETER IDSTEIN, SEYMOUR YOLLES

IN-SERVICE training institutes for science teachers are assumed to have a significant impact on the achievement of pre-college students instructed by the teacher-participants. Usually the institute director is not in a position to evaluate the results of his institute through the controlled observation of student achievement criteria. Even though "for many years evaluative studies have produced data useful for making decisions about curricula and instruction,"² the planners and directors of such institutes have maintained that their objectives were to improve the subject matter backgrounds of their teacher-participants or to introduce modern courses, for example, CHEMS, CBA, PSSC, BSCS.

This paper includes the results of a student-achievement-based comparison of teachers who attended a special National Science Foundation Summer and In-Service Institute in modern chemistry. The results raise questions about the immediate impact

on such participants and suggest that uncontrollable variables may have greater influence on findings than the experimental treatment. Suggestions are made for designing more powerful evaluation strategies which focus on improved pre-college student achievement.

Experimental Design

A nonequivalent control group design, "self-selected" Design 10,³ in which the control group and the experimental group did not have pre-experimental random assignment, was used. The groups were naturally assembled collectives, with some taught by subjects from each treatment. The experimental

³ Donald T. Campbell and Julian C. Stanley. "Experimental and Quasi Experimental Designs for Research on Teaching." Chapter 5. In: N. L. Gage, editor. *Handbook of Research on Teaching*. Chicago: Rand McNally & Company, 1963. p. 220.

* Robert L. Uffelman, Associate Professor of Education, and A. Jon Magoon, Assistant Professor of Education, both at the University of Delaware, Newark; Peter Idstein, Assistant Professor of Education, Washington College, Chestertown, Maryland; and Seymour Yolles, Professor of Education, University of Delaware, Newark

¹ This study was conducted in cooperation with the Delaware State Department of Public Instruction with funds provided by the National Science Foundation and the DuPont Company.

² M. C. Wittrock. *The Evaluation of Instruction: Issues and Problems*. New York: Holt, Rinehart and Winston, Inc., 1969. p. vii.

STUDY IN CUERNAVACA, MEXICO

THE ONE
EFFECTIVE
WAY TO

LEARN SPANISH

STUDY SIX HOURS EACH DAY

FOUR STUDENTS PER TEACHER

PAY \$120 FOR EACH MONTH

START ANY FIRST MONDAY

If you have decided to master Spanish, and if you want to do so in no more than three months, come to CIDOC. While here, you have the opportunity to join informally an academy addressing itself to radical alternatives in our institutions, especially in Latin America. You can also enroll in one-month formal seminars at \$30. Yearly registration fee is \$50. Room and board in Cuernavaca from about \$80 monthly.

For catalog and further information, write:

CIDOC APDO. 479, CUERNAVACA, MEXICO

subjects deliberately sought out the in-service institutes, but no control group was available from this population of seekers. In this case, the assumption of uniform regression becomes less likely, and selection maturation interactions become more probable.

The study did provide information which could in many instances help rule out the null hypothesis, that is, that attendance at the institute had no effect. The design was strengthened in that random samples were selected from volunteers who participated and from volunteers who did not participate. The use of covariance adjustments produces stronger evidence than simple matching. Despite this care and attention, the use of Design 10 requires cautious interpretation of the findings.

The hypothesis tested was: There will be a greater gain in mean scores on the Cooperative Test in Chemistry, 1967 Form, for students whose teachers attended an in-service institute program than for students whose teachers did not attend.

Selection of Subjects

The experimental group consisted of five teachers from different type schools and geographic areas selected randomly from among the institute participants. It included one-sixth of all participants, or one-third of the chemistry teachers, in the institute. A similar sample from teachers of chemistry who had never participated in NSF or other in-service programs was selected as a control group from schools judged comparable to the participant-sample. All teachers had taught a minimum of three years and were using comparable textbooks^{4,5} and laboratory facilities. Classes using CHEM Study⁶ materials were excluded, because studies^{7,8} indicate different achievement results are obtained with this program. Equal size samples were achieved by randomly excluding students until both groups were of equal size (total N=502).

The Criterion Instrument

The instrument used in this investigation was the ACS-NSTA Examination in High School Chemistry, Form 1967.⁹ This instrument is designed to test students who have had some stress placed on learning of skills, principles, and concepts. It contains 80 items arranged in two parts. Either part samples the entire course, is equally valid, and is of comparable difficulty. The use of both parts takes 80 minutes or two test sessions and

⁴ H. C. Metcalfe, J. E. Williams, and J. F. Cafska. *Modern Chemistry*. New York: Holt, Rinehart and Winston, Inc., 1966.

⁵ P. J. Boylan. *Elements of Chemistry*. Boston: Allyn and Bacon, Inc., 1963.

⁶ See: George C. Pimental, editor. *Chemistry, An Experimental Science*. (Chemical Education Material Study.) San Francisco: W. H. Freeman Company, 1963.

⁷ Verne A. Troxel. "Comparison of Achievement of Students in High School Courses." *Journal of Chemical Education* 47 (1): 79-81; January 1970.

⁸ John F. Schaff and Paul H. Westmeyer. "Comparison of Students in Modern and Traditional High School Courses." *Journal of Chemical Education* 47 (1): 82-84; January 1970.

⁹ ACS-NSTA Examination in High School Chemistry, Form 1967. Tampa, Florida: University of South Florida, 1966.

increases the reliability by providing a more thorough sampling. A representative group of chemistry teachers judged the test valid for assessing students in the high school courses taught in the sample schools.

Results

Student scores, after correction for possible guessing (Score = right answers - $\frac{1}{4}$ wrong answers), were divided into two groups, participant and nonparticipant student subjects. These two groups formed two levels of a simple analysis of variance design. Pretest measures on each student served as a control variable, and the contribution these made to post-test variation was partialled out in the analysis of covariance. The group means on both the pretest and the post-test were ranked and a Spearman rank-order correlation coefficient with a moderate value of .62 was obtained. This dependence of the post-test scores on pretest scores indicated that covariance procedures were appropriate.

Students of the "participant" teachers scored significantly lower than students of "nonparticipants" on both the pre- and the post-test measures ($F_{pre} = 10.3$, d.f. = 1,500, $p < .01$; $F_{post} = 26.2$, d.f. = 1,500, $p < .01$). The effect of being a nonparticipant's student could not be accounted for by the differences in the pretest scores on the ACS test only, that is, the F statistic for the test of adjusted means was significant at the .01 level. The analyses of covariance results are as follows:

Adjusted group means	Residual mean squares	Degrees of freedom	F
Participant = 22.5	Between = 3935.31	1	7.99 ($p < .01$)
Nonparticipant = 28.2	Within = 492.55	248	

Conclusions

Rejection of the hypothesis does not imply that this institute actually resulted in a detrimental effect on the students. It seems much more plausible to argue that variables not accounted for had much to do with the outcome. Some support may be gained from

Social Studies Education Projects: an ASCD index

Compiled by

Bob L. Taylor

Professor of Education, University of Colorado, Boulder

and

Thomas L. Groom

Supervisor and Instructor, College High School, University of Northern Colorado, Greeley

Developed for the Generation of New Understandings Project

Annotated listing of more than 100 sources for data and materials

Price \$2.00 Stock No. 611-17844 92 Pages

Order from: **Association for Supervision and Curriculum Development, NEA**
1201 Sixteenth Street, N.W., Washington, D. C. 20036

the post-hoc analysis that revealed socioeconomic and cultural differences between the groups were quite strongly related to both initial and final test scores. Rank-order correlations between the proportion of the student body anticipating attendance at college, a variable closely related to the socioeconomic status of the community, and pretest and post-test group means were .72 and .54 respectively.

Another explanation for these "negative" results might be that teachers who volunteer for institutes may accurately perceive themselves as having less background in chemistry. Such original or perceived differences between a self-selected participant group and a nonparticipant group may be very difficult to overcome. However, it was concluded that the results of this study cast doubt on the practicability of Campbell's Design 10 for investigating the impact of in-service institute programs.

Implications

Some recommendation regarding research strategy is relevant. If a random selection of participants is possible, then traditional experimental designs¹⁰ might be appropriate when adequate controls of many variables are possible. If these controls are not available, a time-series design is recommended on condition that an extended period for evaluation is practical.¹¹ This design employs multiple observations before a planned or naturally occurring intervention of a condition or variable. Observations are then taken after the intervention has occurred. Specific use of this design is applicable to the evaluation of in-service institutes or systems approaches to curriculum improvement.

Such studies should start by instituting annual standardized measures of student achievement in the area being investigated to

provide a year-by-year record of pupil achievement for any given teacher or school. Notation is made of any important changes, and the data are examined to determine if student performance changed from previous years or is maintained over an extended period of time.

The nonequivalence of the groups can be compensated for by using trend-analysis strategies. One first uses a simple one-way analysis of variance. Then, each group is treated as a level of a factor. If any significant value for F is found, a Newman-Keuls test can be used to single out the level or group which is different. The "different" group would be examined to generate plausible hypotheses that could be tested further.

Trend-analysis provides for identifying breaks in chronological patterns within the time series. It can be used to compensate for nonequivalence of groups by providing a homogeneity of variance which could strengthen inferences regarding the effects of the intervening variables. Because directors of in-service institutes can rarely utilize true experimental control groups, the time series offers promise for determining the impact of their teacher training programs.

Synopsis

Achievement gains on the ACS-NSTA Examination in High School Chemistry, Form 1967, for two groups of high school students (N = 502) were compared to determine if this procedure would measure the impact of an in-service institute for teachers and their students. Students of "participant" teachers achieved less gain ($p = < .01$) than students of "nonparticipants." These results led to examination of other variables providing an impact on these subjects, that is, percentage of students in academic programs, socioeconomic factors, limitations of teacher selections for institutes, and applications of experimental design to formative evaluation of teacher education programs. A different set of strategies is needed. Time-series design appears to offer an alternative for evaluating curriculum improvement activities. □

¹⁰ Herbert J. Walberg and Wayne W. Welch. "A New Use of Randomization in Experimental Curriculum Evaluation." *The School Review* 75 (4): 309; Winter 1967.

¹¹ John M. Gottman et al. "Design and Analysis of Research Using Time Series." *Psychological Bulletin* 72 (4): 299-306; 1969.

Copyright © 1971 by the Association for Supervision and Curriculum Development. All rights reserved.