

● Research in Review

COLUMN EDITOR: JAMES RATHS

A Problem of Validity in Curriculum Research

HARRY O. HALL
CHARLES DZIUBAN

INTELLIGENT decision making in curriculum development, by and large, can only be obtained through careful experimentation. Unfortunately, experimentation with human subjects, particularly in the public school setting, sometimes forces compromises in a research design which can give spurious results—and thus lead to erroneous conclusions. The description which follows was written as an illustration of a well-planned project that was emasculated because close examination revealed certain design flaws. As a result, serious doubt was cast on the validity of the data.

Background

In the now popular nongraded high school, English is no longer taught in the same way to all students. This is especially true at the lower ability levels, where students generally spend intensive and extended periods on skills practice in reading and writing. Basically, the logic rests on an assumption that school success and language skills are highly interrelated. The idea seems to work in nongraded schools, but will it work in other situations? To answer this

question for the principal of a large metropolitan high school, a special research study was designed to test the effectiveness of a special language skills course for skill deficient sophomores. The following brief rationale was established:

One or more phases of the language art skills of reading, writing, speaking and listening are involved in some way for virtually every subject activity in the modern high school. Students who have not mastered these skills with some degree of proficiency are under a severe learning handicap. By improving language skills the student should perform more effectively in school.

Student Selection

Junior high feeder schools for the senior high school were each notified of a new course offering especially designed for average students who were deficient in language ability, particularly in reading. It was observed that over 400 students met the basic criteria in data taken from a ninth grade county-wide testing program. The junior high school counselors were apprised of the course as were the English teachers. Both groups were requested to nominate students for the new course.

In an attempt to control certain aspects of student motivation, all parents whose children were identified either by standardized tests or by school personnel were sent a letter announcing the course and requesting permission to assign the child to the experimental course. The student selection technique was to be: (a) identify a large pool of eligible students; (b) select student pairs matched on the criteria of IQ score, reading score, sex and vocational intent; and (c) by a random means, select one pair member for the experimental class and the other for the control.

Fewer than fifty parents responded to the permission request and of this number four were refusals. This disappointing response forced the motivation control to be abandoned. It was assumed that those students without letters did not object to the special course. Only the four refusals were excluded from the study.

As student selection proceeded it was found that those in band, chorus, orchestra, athletics and other subject areas could not enroll in the special class. Since high school students have the option of selecting their majors, little could be done to force anyone into the experiment. As a result, the experimental group was chosen on the basis of those who qualified and who could fit the course into their schedules.

After the experimental group was selected, the control students were chosen on the basis of how well an individual matched another in the experimental class. Matching on the four criteria in the two groups was nearly perfect. Vocational intent and sex in each pair were in agreement and the other criteria varied no more than four test score points.

The mean pair score differences for both reading and IQ were nil.

To prepare for the course a team of three teachers devoted two weeks in a preschool workshop to writing objectives, developing materials and discussing organizational strategy. For a full school year these teachers taught the experimental course. Their 75 students devoted four 25-minute modules of the school day to the language skills course which consisted of instruction and drill in reading, writing, listening and speaking. Students scheduled the experimental class in place of regular English 10 and study hall. On the basis of standardized tests and informal inventories, students were grouped in each skill and received special instruction commensurate with their skill level. All grouping was tentative and based on diagnosed deficiency level.

As a student improved or demonstrated proficiency in a skill area he was moved to more challenging work. Instructional activities varied from individual independent study to total group lectures and demonstration. At the end of one year the results were impressive. A comparison of the experimental class performance with a control group on standardized test performance in reading, writing and listening showed that all aspects favored the experimental class. The analysis of covariance technique using the pretest scores as a control was used to show that a significant difference existed between the means at the five percent level of confidence for the three tests.

In addition, the experimental students displayed promising gains in grade point averages. When ninth year grade point

(Continued on page 89)

(Continued from page 82)

averages for both groups were placed in an array and when the relative standings were compared with the tenth grade averages, significant shifts favoring the experimental class were found. However, the outstanding result was the dropout comparison rate of eleven to one favoring the experimental group.

Internal and External Validity

According to the results of the study, was the principal justified in changing the school curriculum by including the language skills course? Without further investigation, the answer was obvious, but intelligent action dictated studying two important problems associated with all research—internal and external validity.

Internal validity in the reported study was addressed to the question, "Can the observed differences between the control and experimental groups be attributed to the effects of the language class?" With a "yes" answer, it could have been reasonably concluded that language skills training was a good idea for the experimental subjects. Further, if it was concluded that similarly skill deficient high school sophomores profited by taking the special course, then the issue of external validity was satisfied.

Discussion

It should be noted that all of the results reported were comparisons of one group with the other. Consequently, the favorable growth shown by the experimental class was relevant and had value only to the degree that the test groups were evenly matched at the start. Were they, in fact, well-matched on all relevant criteria?

There was definitely a good match on test scores. Since IQ and reading scores were criteria used for pairing, reasonable control was assured. The final results taken from comparable forms of the STEP Tests published by Educational Testing Service in reading, writing and listening were considered reasonable and valid. However, it should be kept in mind that regular English classes teach language skills, but not to the degree that was taught in the experimental class.

For the special class not to excel in language skills would have reflected a serious curriculum design error. Although skill growth was a necessary ingredient in the results, it was not sufficient to sustain the case for the new course. The real point, previously expressed in the rationale, was "would language skill improvement influence the student's school performance?" Many factors affected school performance and probably the most important was the student's motivation to excel in school.

Frymier's *Junior Index of Motivation Scale* gave data indicating that no differences existed. However, other unobtrusive measures were available. The two most obvious and available were ninth-year grade averages and school attendance. No major trend was found in ninth-year grade averages to indicate an imbalance in the two groups. The means for the experimental and control groups were both slightly below a "C" as were the median grade averages. However, in both cases the differences favored the experimental group.

School attendance was another matter. The control group averaged four more days absent than the experimental

group (12 days to 8 days) while the median had a similar trend (9 days to 4 days). Comparing the ninth-year absences for the two distributions showed the experimental highly skewed to low absences while the control was the opposite. Furthermore, the percentage of dropouts for the control group far exceeded the percentage for the original selection pool of 400 students. These were quite significant data since high absence rate has long been known to be a strong correlate with school dropouts. The implication was clear. A strong bias was operating in the groups—strong enough to affect the outcome.

What went wrong? The best guess was that a systematic bias was induced by the selection procedure. When the letter to the parents failed to solicit high interest in the project, it was ignored. This, coupled with scheduling difficulties, led to the experimental group being identified by self-selection rather than by a purely random means. Interestingly enough, the problem would not have been corrected by extending the matching criteria to include days absent. As a practical matter, providing additional criteria, although resulting in more precision, makes matching more difficult. Besides, control by matching will always be limited to known variables. Where certain variables are known to affect outcomes, controls should be applied, if possible. However, random subject selection is the *sine qua non*. With it,

equalization of groups can be claimed for known and unknown factors. Furthermore, randomization is a basic assumption for most mathematical statistics to be valid.

The question of external validity or generalizing the results for the reported study would be somewhat academic since internal validity could not be claimed, but yearly follow-up studies were conducted and might provide information for other studies.

Junior-year data revealed that extreme regression occurred with the experimental students. In the standardized test results, experimental class mean scores fell below the control class in all areas, grade averages dropped, while the control grade averages rose and experimental dropouts increased at an alarming rate.

Regression effects were expected, but the severity of the reaction serves to indicate that the experimental students did indeed benefit from taking the new class. To be sure, what the effect was cannot be ascertained by the reported data, but there was sufficient evidence to encourage further study. Unfortunately, the payoff in the final research report can contain no more.

—HARRY O. HALL, *Professor of Education, Florida Technological University, Orlando, Florida*; and CHARLES DZIUBAN, *Research Assistant, Public Schools, Atlanta, Georgia*.



Copyright © 1967 by the Association for Supervision and Curriculum Development. All rights reserved.