The Appropriate Experimental Unit

JAMES RATHS

ONE idea regarding the methodology of research that is fairly widely accepted among the expert researchers in education deals with the appropriate unit of analysis in educational research (6). Many of us in the past have used the student as the unit of analysis in studies comparing one method with another. The scores of individual students have appeared as the basic data of our statistical analysis. Generally, this approach is incorrect.

This issue is discussed anew in Stanley's latest contribution, *Improving Experimental Design and Statistical Analysis* (4). The discussants in this report of a Phi Delta Kappa symposium make the point that if a researcher tests the hypothesis that Method A is more effective than Method B, and if he assigns one group of 30 students to Method A and another group of 30 students to Method B, the appropriate number to use for the N of this experiment is 2 and not 60. Thus, it is argued that *the appropriate unit of analysis is the classroom and not an individual student.* This position can be defended both logically and empirically.

Logical Support

A unit of an experiment is defined as the smallest entity within an experiment that may receive different treatments (1). All the children in a classroom presumably receive the same treatment—therefore the child as such does not meet the definition of an experimental unit while the classroom unit does. Even if two classrooms are both assigned to Method A, clearly the students in these two classes receive different treatments—namely Method A modified by the interests, skills, and personality of the various teachers. If the same teacher were involved, changes in time of day, constituency of the class, etc., would still make the treatments different. Therefore, the classroom meets the definition of an experimental unit.

Another argument is that children within a classroom affect each other's learning. For example, an extremely bright student may ask questions, clarify murky concepts and otherwise behave in ways that assist all students in the class. As a counter-example, in another class, a student may ask dull

questions, annoy the teacher, or interfere with instruction. What these examples indicate is that learners in a classroom may affect one another. Their achievement scores are not independent. (Supposedly, it is for this reason that the height of a single plant of corn is not used in agricultural research as a unit of analysis because individual plants neighboring a particularly tall stalk are systematically affected by growing in its shade.) The argument continues that since the student scores are not independent, then their variability is smaller than if they were independent. In other words two completely different individuals become less variable when they are exposed to the same experience.

If the variability of the group is thus diminished by common experiences, then statistical tests of the mean differences are inappropriate since these tests assume independence of the data. Unfortunately, the violation of the independence function works to yield more significant results than should be expected by chance, so researchers who use individuals and not classrooms as units in methodology studies generally report spurious significant results associated with their analysis. (This problem of appropriate units is not encountered by researchers who use treatments which present material to students one at a time—such as programmed learning. However, if the programmed learning sessions include group discussions, then the question of appropriate unit is still relevant.)

Empirical Support

Of course all of the above is merely a logical analysis of what could happen. Is it not possible that in every classroom, for every student who helps other students by his behavior, there is a second student who acts in ways to inhibit learning with a resultant zero effect?

Steck (5) tested the statistical hypothesis that the scores of students learning in a group have a smaller variance than do the scores of students learning on a one-to-one basis. Thirty 7th grade students were randomly assigned to receive instruction in mathematics on a one-to-one basis while another thirty received instruction in a group. The presentation of the lesson was made by tape-recorder to ensure similarity of treatment. In both situations, students were encouraged to ask questions concerning the lesson before they took an achievement test to ascertain the extent to which they had mastered the material.

Steck found that the variance of the scores of students who learned in a group were significantly smaller than the variance of the scores of students who had learned individually. He concluded that these data supported the hypothesis that common experiences in a learning situation reduce the variances of student scores.

Implications

It is hoped that graduate student advisors, editors of research journals, and consumers of research will hold in disrepute all studies which attempt to use individual students as the unit of analysis in a methodology study not involving a treatment presented to individuals.

One source of misunderstanding on this score may well come from the examples found in most statistics books.

So often, a Method A *vs.* Method B setting is suggested in the exercises and students are encouraged to apply a t-test, for instance, to individual student scores. Authors of statistics books could give emphasis to the appropriate unit of analysis in their presentations.

Finally, the situation does not necessarily require researchers interested in testing a Method A *vs.* Method B hypothesis to find thirty classes to receive Method A and thirty classes to receive Method B, although this is certainly an appropriate alternative. (In the analysis, the mean scores of each group would become the criterion measure and the analysis would be carried out with an N of 60.)

Page (3) has offered another alternative that requires an analysis a bit more complicated than a simple t-test, but one that can be rather easily applied. Briefly, Page advocates that each classroom should be treated as if it were a single subject and then treat various sub-categories found within the classroom—e.g., sex, levels of intelligence, levels of achievement, etc., as though they represent measurements of the same subject made under different conditions. This approach allows a researcher to use the designs of repeated measures found in Winer in his chapters 4 and 7 to complete the analysis of the data.

This design overcomes the logical problems associated with the violation of assumptions encountered in the application of a t-test where it is inappropriate. Yet, as Lumsdaine (2) pointed out in a brilliant and generally overlooked chapter on research in the well known *Handbook for Research on Teaching* (2), such a design not only takes into account "the fluctuations due to individual-difference (population) variables, but also properly reflects the influence of variations for different classroom groups, e.g., time of day, instructor, behavior, procedural irregularities, *etc.*" (p 657). Hopefully, more and more curriculum researchers will adopt the more rigorous design advocated by Page so that more reliable findings may be utilized in making curriculum decisions.

References

1. D. R. Cox. *Planning of Experiments.* New York: John Wiley & Sons, Inc., 1958. p. 24.

2. A. A. Lumsdaine. "Instruments and Media of Instruction." *Handbook for Research on Teaching.* N. L. Gage, editor. Chicago: Rand McNally & Company, 1963.

3. Ellis B. Page. "Recapturing the Richness Within the Classroom." Paper presented at the American Educational Research Association, Annual Meeting, Chicago, Illinois, February 11, 1965.

4. Julian Stanley, editor. *Improving Experimental Design and Statistical Analysis.* Chicago: Rand McNally & Company, 1967. p. 183.

5. J. C. Steck. "The Independence of Observations Obtained in Classroom Research." Unpublished Master of Arts thesis. College Park: University of Maryland, 1966.

6. David E. Wiley. "Standard Experimental Designs and Experimentation Under School Conditions." Paper presented at the American Educational Research Association, Annual Meeting, Chicago, Illinois, February 11, 1965.

7. B. J. Winer. *Statistical Principles in Experimental Design.* New York: McGraw-Hill Book Company, Inc., 1962.

—JAMES RATHS, *Director, Bureau of Educational Research and Field Services, College of Education, University of Maryland, College Park.*