

ences needed by these particular individuals to accomplish the desired ends? This will depend in large measure upon the nature of their previous experiences, and here there is apt to be a great deal of variation. Again, there is no more valid data than those which can be provided by the class itself. With such data at hand the teacher is in a stronger position to decide whether certain types of experience are genuinely needed by all and whether other types are best used for particular individuals and small groups. Probably one of the greatest wastes of effort in our whole educational enterprise comes from the trouble to which we go to insist that all students have experiences which may be repetitive for some, and hence inappropriate, or which may be

premature for others, and hence likewise inappropriate.

Since the matter of how the classroom teacher gets and uses evidence is treated elsewhere in this issue, the point need not be labored here. It is perhaps sufficient to remind ourselves that decision-making in the classroom is the decision-making that counts. It determines what the actual instructional program will be. It provides the conditions which are thought to be appropriate in order to have genuinely effective instruction. With so much at stake, it would seem that the conscientious teacher has no alternative but to accumulate continuously and systematically the kind of evidence he needs to guide his own decisions with respect to instructional problems.

JULIAN C. STANLEY

The Interdependent Roles of Research and Evaluation in Teaching

"The teacher introduces new procedures after analyzing their relevance for his situation and then evaluates their impact upon the students with respect to objectives of instruction."

ROUGHLY, the somewhat vague word, "research," refers to three interrelated types of activities: consideration of written and oral reports by persons other than oneself, experimentation, and controlled experimentation. When speaking of the "experimental method" in education, frequently we mean trying out new approaches rather informally in the classroom, without most of the controls characterizing a labora-

tory experiment, and attempting to assess the effectiveness of these procedures by synthesizing various types of information subjectively. Such experimentation may stem from library research, discussion of teaching methods with others, in-service training and courses taken, urging of supervisors, and systematic theory. Experimentation starts from the same bases, especially library research and theory, but

Julian C. Stanley is associate professor of education, University of Wisconsin, Madison. Dr. Stanley is at present on leave, as a postdoctoral fellow in statistics and education at the University of Chicago until September 1, 1956.

by careful advance planning (called experimental design) it maximizes the objectivity of assessment.

Educators endow the terms, "measurement," "evaluation," and "assessment" with various meanings. The last two imply objectives of instruction, well thought out in advance, with carefully prepared, varied and frequent checks upon their attainment. Teachers and specialized personnel attempt to use inventories, self-report devices, projective techniques, questionnaires, rating scales, sociometric procedures, anecdotal records, conferences, recordings, and films in optimum combinations for appraisal of the "whole child" and the entire educational setting.

Evaluation, a more inclusive term than measurement, connotes the making of value judgments, particularly in rating the desirability of educational objectives. Some writers reserve evaluation for this limited use, calling the procedures by which we appraise the growth of individuals with respect to these objectives, "assessment." But in the rest of this article let us adhere to the customary definition that subsumes measurement under evaluation, allocating intuitive, informal judgments of the pupil's progress and aspects of valuing to the latter. In other words, we would call the more objective aspects of evaluation, especially the use of "ability tests," measurement, and the more subjective and intuitive aspects, such as anecdotal records,

evaluation; all measurement constitutes evaluation, but not all evaluation should be referred to as measurement. Admittedly, we make this distinction for convenience in the subsequent discussion, though quite properly you may prefer other usages.

Now that we have informal definitions of research and of evaluation, let us look at the interdependency of these two processes. As indicated above, teachers continually modify their methods in order to maximize the probability of attaining desired objectives and try to evaluate pupil growth to determine whether or not the new procedures facilitate this attainment. Clearly, teachers should have reasons other than caprice or desperation for changing their methods, and techniques other than prejudice for evaluating the impact of modifications. Appropriately, the teacher prejudices a method as being feasible and consonant with educational objectives before adopting it, but this value judgment does not insure the method's efficiency. Only periodic evaluation, planned from the beginning so as to minimize prejudice, suffices to determine whether the adopted procedure "works."

In a loose sense, all non-random changes in the treatment of students *result from* research, at least the subjective analysis of one's own experiences, and *constitute* research, for each change rests upon a tentative hypothesis (rather than an assumption) that it will facilitate the attainment of certain educational objectives. The better the modifications fit into a coherent educational philosophy and the more they stem from well controlled research, the greater the likelihood

that they will expedite pupil growth. However, one must not dispense with evaluation, no matter how promising the new method appears. "On-going" evaluation keeps education "dynamic," protecting it from perpetuating errors of judgment that inevitably occur in an area where the probability of making correct decisions lies well below 100 per cent.

Thus far we have confined the discussion to "methods" research—the instituting of changes and the evaluation of their influence upon the achievement of desired objectives. If this type of investigation is well rooted in prior research and systematic theory and includes suitable controls, it can make original contributions. Broadly speaking, we may classify research as suggesting hypotheses, testing *ad hoc* hypotheses (relevant to a specific problem, but apparently not contributing to theory), or testing hypotheses with theoretical implications. As presently conducted, most "research" by teachers at best generates hypotheses, rather than tests them; a little of it involves carefully attacking special problems suggested not by relevant theory but by "practical" exigencies such as the conviction that certain current practices are ineffective. In the opinion of the writer, controlled experimentation firmly grounded in theory and contributing to it receives infinitesimal attention below the college level. Surely, theory-oriented controlled experimentation in the classroom deserves hundreds of times as much support as it now gets.

Three Planning Situations

Let us look at three illustrations of

informal "research" from among the many that might be considered. Later we shall outline a more structured research project.

● Mr. X, supervisor of social studies in a large school system, reads in a professional journal Professor Y's apparently conclusive evidence of change to more democratic attitudes in small groups of male college sophomores following certain procedures. Mr. X and his teachers note the kind of prejudice among some of their students that the professor attacked successfully in the laboratory. They spiritedly discuss the applicability of his principles and procedures to their situations. They consider performing modifications of his experiment in their classrooms but conclude that limitations of time and resources forbid. Therefore, they read about other investigations, particularly those cited by Professor X, and try to reach consensus concerning seemingly applicable procedures. Finally, and quite importantly, they carefully explore possibilities for initial and on-going evaluation of the new methods.

● A group of high school mathematics teachers meets several times to discuss implications of the NEA's monograph, *What Research Says to the Teacher: Teaching High-School Mathematics* (2). Individuals consult some of the 15 references cited and share their findings with one another. Perhaps no general reorientation of teaching methods evolves, but each person gets certain ideas to try out rather informally. He makes these explicit, both in writing and orally, so that others can help him plan the evaluation implied by the notion of

"trying out" an idea. He realizes that continual evaluation need not be haphazard and ultra-subjective if it is incorporated into the plan for curriculum change.

● A few elementary grade teachers meet to devise an objective test of the common unit they have almost completed. In the process of discussing aims and objectives, relative emphases, and level of generality of items, they learn much about teaching procedures and decide henceforth to plan the evaluation of their units at the inception rather than only at the end. Nevertheless, they now construct a test, administer it, perform an item analysis (3), and get together again for post-mortem consideration of implications. They infer a great deal concerning mental processes of the students from careful consideration of responses to the various item options and thereby further modify their ideas about content and methods. (Of course, these teachers know full well that objective tests alone do not nearly suffice for evaluation, even of strictly "factual" material.)

The above paragraphs illustrate the interdependency of research utilization and evaluation. As already implied, both "research" and "evaluation" denote various things, from the controlled experiment of the physical scientist and the height measurement of the anthropometrician to the subjective day-by-day experience analysis of the classroom teacher, which some consider research, others evaluation, and many neither. We hear of "action," "service," "cooperative," "flexible," and "on-going" research, the meanings of these adjectives varying

from report to report. Often, research by teachers involves getting facts, suggestions, and ideas from written and oral sources, relating these to the teachers' own methods and philosophies, trying new classroom approaches, evaluating their effectiveness, modifying procedures, and re-evaluating.

Experimentation

In keeping with prevalent educational philosophy, teachers continually evaluate on the basis of their experience (which includes knowledge of new research) both objectives and means for attaining these. Objectives must be consonant with the capabilities of the child. Would any teacher conclude that, because of the current shortage of scientists, an objective for all students should be mastery of the differential and integral calculus by the end of the twelfth grade? Likewise, means depend upon philosophical considerations as well as upon the findings of educational psychology and other disciplines; an efficient method may be philosophically unacceptable.

Perceptive, intuitive teachers, unawed by the "authority" of experts, modify and assimilate research conclusions in an unstereotyped manner, incorporating certain elements without prejudice but tentatively, subject to continued evaluation. Many educators hope that controlled experimentation under classroom conditions will increase greatly, but few feel optimistic, especially because of the rising tide of enrollment, with concomitant shortages of staff and resources. Meanwhile, most teachers must borrow research findings, "experiment" intelligently, and evaluate incessantly. From the

expertly analyzed blend of outside material, classroom manipulation and observation, and systematic evaluation should come improved educational experiences for our children.

Severe difficulties in evaluating the influence of new procedures based upon outside research occur, largely because the absence of concurrent control groups confounds the effects of methods with temporal influences. For example, if the present class knows considerably more English grammar at the end of three months than last year's class did, should this difference be ascribed to new techniques, to initial superiority of the current students, to enhanced learning conditions not directly connected with the classroom changes, or to other factors unaccounted for? By using last year's class as the control group, we create this dilemma. While never easy, proving "causation" is especially difficult when suitable controls are absent. To some extent, proper evaluation can substitute for a contemporary control group, as when the teacher knows from tests how the initial knowledge of students this year compares with that of last year's class and how gains of the two groups differ. Even such seemingly intangible aspects as classroom "atmosphere" and morale can be evaluated with moderate success (1). The teacher strives to objectify and quantify evaluation procedures without narrowing their scope unduly; not all subjective judgment can or should be eliminated, but usually its minimization improves over-all evaluation.

Evaluating a New Method

Recently the writer was privileged

to help plan the evaluation of a correctional speech procedure in the seventh grade classes of a university laboratory school. On the basis of their wide experiences, three speech teachers and the principal decided to have student teachers observe unobtrusively the speech of seventh graders in non-speech classes and give the youngsters written suggestions for improving their diction, volume, pitch, etc. The usual approach might have been to observe all seventh graders, furnish suggestions to those who needed these, and finally evaluate by having staff, student teachers, and perhaps students and/or parents report how well they felt the innovation had "worked."

Instead of adopting the above haphazard evaluation of outcomes, obviously complicated by maturational influences, we set up controls deemed sufficient to make evaluations of outcomes relatively unambiguous. Taking two factors (reported to vs. not reported to, took English first semester vs. took speech first semester) into account separately for boys and girls yielded four *random* groups for each sex: reported to, took English; not reported to, took English; reported to, took speech; not reported to, took speech. At the end of the first semester, each of the students appeared (in random order) to be rated on two aspects of speech by judges who did not know any student's classification. This study continued through the year, those who took English the first semester having speech the second, and vice versa, with intermediate and final evaluations.

The high confidence and enthusiasm of the four persons who devised the

plan for observing the speech of students in their natural habitats and commenting about it conflicted sharply with the negative conclusions of the experiment. Very likely, informal evaluation would not have been precise enough to override initial glowing expectations; one does not want to bury his favorite brain-child without conclusive proof of death.

The structured evaluation described above conserved possible values of the innovation for 50 per cent of the students. By withholding reporting from half of the seventh graders, we obtained data essential for making several crucial comparisons. If the innovation had proved effective, these control students would have deserved extra attention after the experiment. However, they served to show that the considerable expenditure of time and staff required for the reporting procedure *as currently conceived* could be saved henceforth. Certain modifications might produce better results; the speech teachers can explore these on the basis of professional literature and further experimentation.

Research and evaluation go together, being at times virtually interchangeable or even synonymous. The teacher introduces new procedures after analyzing their relevance for his situation and then evaluates their impact upon the students with respect to objectives of instruction. He strives to eliminate many subjective elements from evaluation, while realizing that personal judgment is essential for the synthesis of information and the making of decisions. Ideally, evaluation becomes increasingly objective and quantified, moving toward controlled experimentation.

References

1. Cornell, Francis G.; Lindvall, Carl M.; and Saupe, Joe L. "An Exploratory Measurement of Individualities of Schools and Classrooms." *University of Illinois Bulletin* No. 75; 1953.
2. Fehr, Howard F. "Teaching High-School Mathematics." *What Research Says to the Teacher*, No. 9. Washington, D. C.: National Education Association, 1201 Sixteenth St., N. W. 1955. 25¢.
3. Ross, C. C., and Stanley, Julian C. *Measurement in Today's Schools* (3rd ed.). New York: Prentice-Hall; 1954. See especially p. 436-53 and 117-19.

Announcing

Three New ASCD Booklets:

- Reporting Is Communicating: An Approach to Evaluation and Marking.** Written by the Texas ASCD. (64 p.) \$1.00
- Selected Bibliography for Curriculum Workers (1956 Edition)**
 Edited by Wanda Robertson. (64 p.) \$1.00
- Curriculum Materials for Creative Thinking, Living, Teaching.**
 A catalogue of the Curriculum Materials Exhibit at the 1956 ASCD Conference. (40 p.) \$.75

order from: ASCD, 1201 16th Street, N.W.
 Washington, D. C.

Copyright © 1956 by the Association for Supervision and Curriculum Development. All rights reserved.