

ments can be made available to school people and others.

This year we will give renewed attention to controversial issues. As examples of this interest, we suggest that you read with care the article by Anna K. Lehnhoff and the "essay in criticism" by Alexander Frazier. Do you agree with the viewpoints expressed? Do you disagree? Why not try writing a letter to the editor giving your reactions? We need to increase and sharpen our "response from the audience." We need to know what the audience is hearing, and to what the audience is reacting.

We will again carry a number of "unsolicited" articles. These manuscripts are chosen from among the many that come to us for possible use. We are highly selective with these materials.

Another area of emphasis this year is that of book reviews. Now under direct supervision of the journal editor, the review column will be extended somewhat, so as to include more titles. Reviewers will also try more "combined" or "joint" reviews, treating two or more new books dealing with related topics.

Again, we do not know where the year will take us qualitatively. We believe, however, as we always do at this point in starting a new volume and a new publication year, that each issue will grow into its own demonstration of worth and merit. We believe that each issue will rank well with all preceding numbers of this professional journal.

—ROBERT R. LEEPER, *Editor*, EDUCATIONAL LEADERSHIP; *Associate Secretary*, ASCD.

Evaluation and Curriculum Development

THE interaction between evaluation and curriculum development is intimate and total. Changes can legitimately be made in an instructional program only when careful evaluation demonstrates the strengths and weaknesses of such adjustments. Conversely, no curricular proposal can claim widespread support until and unless it has justified itself through carefully collected data. Evaluation may be called the other side of the coin of curriculum development.

Tests constitute the principal tool of those who evaluate, and indeed of all who would contribute to educational improvement. The testing movement represents a major breakthrough in the advance of education toward the status of a science: without accurate measures there can be no science. The achievements of test makers in quantifying elements of the human mind and personal-

ity have done much to multiply and extend the effectiveness of the teacher.

Consider a school without access to information about intelligence, aptitudes, interests and achievement. Grouping would be done on a catch-as-catch-can basis, counseling would be largely guesswork, individualization of instruction would be based upon doubtful evidence.

However, the testing movement can constitute just as much a barricade to educational advance as it can provide a breakthrough, mainly because the act of evaluation is so charged with emotionality. During examinations and at report times nervous tensions may pervade a classroom like a fog. And teachers, facing evaluation in their own turn, whether from supervisors or college professors, usually view the process with anxiety. All the while the person who evaluates stands alone, filled with self-doubt, as he

realizes his decisions present a picture of omniscience which is more apparent than real.

It is paradoxical that testing and evaluation present a simultaneous picture of objectivity and subjectivity. In no area is the goal of scientific preciseness sought with greater intensity than in that of educational measurement; test technicians struggle to increase the validity and reliability of their tools, using elaborate statistical analyses to this end. And yet the fear and insecurity with which educators approach the task of measurement go far to frustrate these efforts and to block educational progress.

The Difficulties

There are at least four problems which have resulted from the measurement movement, and especially from the development of standardized tests.

1. *There is greater emphasis upon that which is easy to measure than upon that which is important to measure.*

The first successes in any comparatively new field usually take place in those aspects which are most amenable to attack. In educational testing this has meant an early emphasis upon measuring acquisition of facts and basic subject-matter skills. Today, in spite of the many efforts to move beyond such areas, a review of most tests, whether teacher-made or standardized, reveals a continuation of this emphasis.

Concurrently, there is neglect of more elusive and vital learnings such as critical thinking, problem solving, creativity, cultural appreciations, work habits, and the like. Because it has been easy to develop and use measuring devices associated with the simpler learnings, there has been a dangerous tendency to over-

concentrate upon such learnings and ignore more complex ones. Or, perhaps it would be more accurate to say that the emotional hazards of venturing among these controversial and difficult areas have discouraged many persons from making such explorations.

2. *A spurious sense of certainty has come into widespread existence because of uncritical use of standardized instruments.*

The drive for certainty characterizes all human beings. Paradoxically, this drive motivates both adventurers and stay-at-homes. It stimulates many to seek new learnings, among them those persons who have secured for all mankind the vast achievements of research. The same emotion epitomizes those who cling to the security of the past. This dualism partly explains the similar paradox in educational evaluation. One way in which educational stay-at-homes reveal themselves is through their uncritical use of standardized tests.

The trappings of science adorn most test manuals. This remark is not intended to imply hypocrisy or ineptness on the part of test makers, since there is nothing spurious about the efforts of those who create and standardize educational instruments. Indeed, their writings abound with cautions and provisos. Nevertheless, in spite of their careful efforts, or perhaps even because of these efforts, many persons are inclined to place greater reliance upon test results than the instruments warrant. The exactness with which scores are reported, and even the preciseness of the measures of error, delude many people into thinking that a test provides more certainty than it actually does. Further, the question is too rarely raised as to the degree of relevance which seemingly exact measures bear to

the trait or learning they ostensibly describe.

Just because a test uses the words, "Arithmetic Achievement" or "Critical Thinking," in its title does not mean that the definition of these terms is the same in the minds of the test makers as it is in the minds of the test users. Indeed, there is often a question as to whether or not an instrument even reflects the intention of its creators accurately.

3. *There is a de-emphasis on locally constructed measures.*

Educational evaluation is in danger of becoming a spectator sport. The elaborate and costly preparation considered essential to the commercial production of tests has caused many local educators to lose faith in their own devices and skills. While it is true that such persons are generally not highly skilled in techniques of test construction, they do know boys and girls and they do know education. It is important not to discourage these educators from using their own insights in educational evaluation; on the contrary, every effort should be made to encourage and assist their creative endeavors to assess students' learnings.

4. *There has been a notable lack of communication between teachers and test technicians.*

Even as the medical specialist sometimes seems to patronize the general practitioner, so the test specialist often appears convinced that he has been vouchsafed a revelation not granted to lesser beings. Conversely, teachers sometimes act as though they possess an insight into the art of education totally beyond the grasp of those who would capture it in charts and graphs and statistics.

Too often teachers are unduly resistant to the complexities of standardized test-

ing, in spite of its possible benefits. Too often, too, tests are selected and administered with little regard to the reality of the classroom and in a manner actively discouraging teacher experimentation and creativity. Apparently the emotional nature of evaluation creates barriers to communication.

The Challenge

The challenge is to take full advantage of our existent measuring devices and to do so in a way which will encourage rather than retard further educational progress. We must strive to resolve the various areas of emotionality. Following are a few guidelines we might consider.

1. *"Teach for the test," but test all important educational outcomes.*

We have heard many criticisms of the tendency of teachers to "teach for the test." However, such activity is surely defensible if the test (or tests) validly measures all significant student learnings, and if individual needs are recognized. Moreover, it is human nature to place emphasis upon that which will be used to evaluate one's efforts.

Consider, for instance, the difference in a student's study of a book if he is engaged only in self-improvement or if he is preparing for an examination. Teachers will very likely always teach for the test. This fact can be used as an important means of influencing the curriculum. Our job is to be certain of a wide range of valid measuring devices covering all important learnings.

2. *Improve the measurement skills of teachers and the educational skills of test specialists.*

Through joint conferences and in-service education activities, teachers and test
(Continued on page 79)

cepts and attitudes, for improving problem-solving abilities and skill learning, for encouraging creativity and personality integration, and for facilitating motivation, retention and transfer. The statement of each instructional principle begins with a verb; the teacher is the assumed subject. Each principle is well documented with supporting research. The principles are stated in such a way that they will be clear-cut guides for achieving outcomes efficiently.

In view of the inroads of automation and technological developments on the educative process, the chapter on new settings and contexts for learning is a valuable contribution. Descriptions of technological and other innovations are well handled, available evaluative research is discussed, and some provocative research questions are raised.

The entire section on evaluation and measurement, consisting of three chapters, is noteworthy. Particular strong points include the discussion of the specific purposes of evaluating and measuring, the descriptions and illustrations of commonly used standardized and other published tests, and the discussion of research instruments used in evaluating teacher effectiveness.

General weaknesses of *Learning and Human Abilities* can be classified as faults of omission. Although developmental trends of youth are discussed in connection with acquiring educational outcomes, there is no comprehensive treatment of growth and development. For those institutions which include human growth and development in the introductory course in educational psychology, the text would need supplementation. While research studies in education, psychology and other areas are reviewed extensively throughout the text, the motivated student will want to obtain

more from these researches than can be put in an introductory text; namely, the rhetoric of conclusion. For this reason it would be desirable to augment the text with a book of readings in educational psychology. It would be advantageous, also, to supplement the final chapter on statistical and research terminology with a standard introductory statistical reference.

These omissions, however, do not seriously detract from an introductory educational psychology text which should find a warm reception from the profession. This volume should be particularly useful for introductory courses in educational psychology with prospective teachers. All those who are concerned with the efficiency of student learning in schools should find *Learning and Human Abilities* of value.

—Reviewed by RICHARD E. RIPPLE,
Assistant Professor of Educational Psychology, School of Education, Cornell University, Ithaca, New York.

Editorial—Evaluation

(Continued from page 6)

specialists must come to know each other's needs, concerns and insights. Both should possess a veto power over test selection in the individual school, since neither group alone is qualified to choose wisely. If the two cannot agree, it is probably best to avoid choosing any standardized instrument within the area of disagreement but to use locally developed devices instead. Further, such lack of agreement indicates a possible focus for in-service education.

3. *Make teacher-made tests an integral part of the total evaluation program.*

It is a travesty to call a program "The Testing Program" of a school when this program contains only a minor fraction of

the tests used. The bulk of educational measurement and evaluation is done in the classroom by the teacher, and must be harmonized with findings of "The Testing Program." An important part of this effort should be the continuous provision of specialized assistance to teachers in developing and improving their own instruments.

Planning an evaluation program and creating measurement devices are as much an exercise in educational philosophy as they are a task for the statistician. The validation of instruments and procedures requires coordinated efforts of everyone, since the task involves educational goals, classroom practices, conditions of learning, characteristics of children, as well as criteria of test construction. For too long a period, we have acted as though all this could be done without the participation of those who know curriculum and instruction best. It is clearly evident that this is not the case, and that the improvement of testing and evaluation in schools is interlocked with the improvement of instruction. Perhaps it might be said that the basic job is one of finding ways to quantify and communicate the observations, judgments and experiences of teachers.

—CLIFFORD BEBELL, *Director, Elementary and Secondary Education, State Department of Education, Denver, Colorado.*

Teacher-Made Tests

(Continued from page 19)

making the stated objectives of a course or unit more meaningful for all the teachers involved. Another dividend ensuing from joint work on test questions comes from seeing how others pose questions requiring the understanding of a concept or relationship. The number of unique questions, so necessary for measuring

understanding, can thereby be expanded and made part of a continuous test file available to all teachers.²

Although competent teachers are the best judges of the correctness of a given response to a question and the incorrectness of the remaining responses, any teacher who discusses questions with his students after a test has been given soon becomes aware of ambiguities undreamed of earlier. One way of systematically checking test questions for clarity and accuracy is by means of an item analysis. This involves a comparison of approximately the top quarter of the papers on a test with the bottom quarter of the papers. If fewer or even the same number of students in the upper group compared to the lower group respond incorrectly to a given question, that question may well need revision. By noting which incorrect response is attracting the better students, it is possible to identify the difficulty.³

A teacher who takes the time to clarify the major concepts and relationships he desires his students to understand, follows this up with questions requiring understanding, and then revises questions to improve their clarity and correctness, has made testing a positive rather than a negative factor in the educational process. Then, even those students who study only to do well on a test will also be improving their understanding of the objectives the teacher considers important.

² A file of social studies test questions can be expanded immeasurably by obtaining several bulletins, each containing several hundred questions, published by the National Council for the Social Studies, 1201 16th Street, N.W., Washington 6, D.C. Bulletins now prepared deal with Study Skills and Critical Thinking (#15), World History (#9), American Government (#13) and American History (#6).

³ A clear, brief description of how to perform an item analysis is provided in "Short-Cut Statistics for Teacher-Made Tests," Evaluation and Advisory Service Series No. 5, Educational Testing Service, Princeton, New Jersey.

Copyright © 1962 by the Association for Supervision and Curriculum Development. All rights reserved.