

All Teachers Make Tests

and in-service workshops can help!

CURRICULUM workers and administrators are often absorbed and concerned with the problem of *deciding what to teach*. This involves much committee work and discussion of broad educational and curricular objectives and of specific subject matter objectives. Curriculum guides and courses of study often spell out these curricular, course level, and unit objectives as well as the content for learning. Only rarely do these guides indicate specifically the learnings that teachers should evaluate.

It is true that many objectives may not be measurable immediately; however, it is also imperative that teachers and curriculum workers assume the task of identifying the objectives, behaviors and content which can be tested. Such identification requires skill and discrimination. It also requires an ability to design test specifications and test questions. This is a task that would consume a great deal of teacher time if approached scientifically. The skills and competencies teachers need to develop test specifications and discriminating test items are often lacking and the time re-

quired to design an effective test is rarely provided by the administration.

How many teachers approach testing by drafting specifications for content, for level of difficulty, for item types and number of items, for time limits? How many teachers pretest items and analyze their effectiveness? How many teachers know that there are technical factors involved in test construction? Do teachers realize that students' marks, which are often based on teacher-made tests, are only as valid or invalid as the tests they have constructed?

A Study of Testing

Due to poor tests, many sins are committed in reporting pupil progress. Teachers must begin to realize that their tests are useful not only for evaluating students and assigning marks, but also as a device for self-evaluation. Testing is an important phase of the teaching-learning process. Testing *can* be teaching; often the format of a test is the determining factor.

We found in the School District of Cheltenham Township that many teachers were aware of their shortcomings in test construction, particularly in their power to develop discriminating test

Frances R. Link is Coordinator of Secondary Education for the Cheltenham Township School District, Elkins Park, Pennsylvania.

items. Social Science and English teachers expressed great interest in a proposal to bring consultants in to work with teachers. Thus the plan for our in-service program developed partly out of a desire on the part of some teachers to become more competent to perform their teaching duties. The plan also grew out of administrative dissatisfaction with the lack of knowledge of the technical aspects of test construction and statistical treatment of test scores which teachers professed in the employment interview.

These factors encouraged the administration to plan a series of in-service workshops to improve teachers' skills and knowledge in what we *do* know about good testing. The Superintendent of Schools, Director of Elementary Education, Coordinator of Secondary Education, and the School Psychologist consulted with the Educational Testing Service in Princeton, New Jersey.

Following the guidelines cooperatively established by the Cheltenham staff and the E.T.S. consultants, we spent a year of planning for an in-service program to improve teacher-made tests. Planning meetings involved key school personnel, including principals, guidance counselors, curriculum committee chairmen and department heads. Then the Evaluation and Advisory Service at E.T.S. worked with the school personnel in designing the program described here.

Eight meetings were planned to treat six general topics and two specific content areas. The general topics concerned:

- Principles of Measurement
- Principles of Test Development
- Item Analysis, Scaling, Combining Test Scores
- Techniques of Essay Testing
- Problems with Respect to Assigning Marks
- Interpretation of Test Results.

Fifty key school personnel involved in the original planning attended six half-day sessions with the consultants who described construction of tests, scoring, data handling, and test interpretation. The relation between the development of teacher-made tests and standardized tests was stressed from the point of view of similarities and differences in purpose for evaluating the learning process in the elementary and secondary schools.

One of the eight meetings was planned for teachers of English Language Arts and another for teachers of the Social Sciences.

The Social Sciences meeting was particularly successful. Dana G. Kurfman was the consultant. He stressed the extreme diversity of social studies curricula from place to place throughout the country, and how racial, religious and economic issues and problems present difficulties for the developers of standardized tests. Some of the same problems confront the teacher. Nevertheless teachers must make their own tests to assure validity in terms of local school and community objectives. As individual teachers or groups prepare test items, each item or question should exemplify some objective. In this way, more focus will be placed on objectives which should result in better teaching.

Some of the many inadequacies which have *not* been dealt with in the social sciences at the curriculum planning level include differences in interpretation and concepts, and distinguishable differences between elementary, junior high, senior high and college courses of study. Greater articulation is needed among the levels in curriculum planning and test development.

The most controversial issue of this in-service program on testing centered

on the objective test versus the essay test. English teachers were particularly vocal on this subject; as a matter of fact they became indignant and emotional in their defensive support of the essay test. Consultants aroused the teachers with some of the following statements:

"The essay test does not measure the students' ability to write more effectively than the objective test."

"There is an honest difference of opinion as to what constitutes good writing."

"Essay tests can not be graded reliably."

"The work of the student in class may color the score of the essay."

One of our consultants, Robert L. Ebel, presented the following viewpoint on essay tests:

Objective tests are designed to obtain maximum reliability in the test taker's performance and in the scorer's evaluation of his performance, with minimum loss in the validity of the test performance as an indication of present achievement or future potential. Most, if not all, of the things which can be measured by written essay tests can be measured more reliably and validly by objective tests. Sometimes, however, essay tests may be more convenient to use (as when the number of persons to be tested is small) or may have more desirable side effects, such as encouragement of the development of writing ability.

Paul R. Diederich, another consultant, described the essay as a powerful instrument having two major problems in its use: one, the infinite evasiveness of students and, another, that the mere use of language can get the sympathy of the instructor. He stressed that teachers need to beware of feeling that essay grades are true; we know they are not. Research indicates that the best reliability among reliable readers grading the same paper is .75. However, essays should be used in all grades and subject areas; it is an

excellent device to help students put facts into order. Dr. Diederich chose to depict the essay as an oil painting and the objective test as a series of candid snapshots.

Teachers stated that they found very useful the suggested criteria for good essay questions and the suggestions for use of the essay which several consultants presented:

1. The question that produces the widest possible range—of "good" and "bad" responses—is the best type of essay question.

2. Poor student writing usually results from assignments about writing. Students learn to write better when a human situation is given to them, e.g., a composition on "Huck's Moment of Decision," not on "Techniques of Satire in *Huckleberry Finn*."

3. A suggested project for student writing was to take a point in time and have students discuss what went wrong and what went right in that particular period. For such a project and for all their writing, students should be held responsible for pertinent information.

4. Writing should encourage students to develop knowledge of the interrelationships of various kinds of knowledge and subject matter.

5. Do not use essay type tests when an objective type will serve the purpose better. An objective test will test spelling, usage and vocabulary; but to test ability to write a composition, only a composition will do.

6. Use as many questions as possible: for a one hour test, four short essays are better than one long essay; for a three hour test, fifteen short essays are better than three or four long ones. The principle is to avoid the type of question which would take an hour to develop.

7. Do not give students a *choice* of essay questions. If you instruct students to choose any two of three questions, the chance of achieving reliable grades is lessened considerably.

8. Before an essay test is administered to students, the teacher should write out a

satisfactory answer for each question, and decide on the exact scoring scheme. (Do not mark for anything except what the question is testing for.)

9. When scoring essay tests, score all papers for a given question throughout.

A useful suggestion made by Dr. Diederich for the improvement of essay grades is to keep tabs continually on how they are working. This can be done by having each department that uses essay grades meet after important tests have been administered, so that all teachers may experience grading the same papers independently. They should then argue out their differences. This could help to provide a periodic check on the reliability of grading and on test content validity.

Thirty-two participants in the in-service workshop completed an anonymous questionnaire to help us to evaluate the worth of the program, and to establish guidelines for future planning. The workshop experience, in general, was rated as follows:

Excellent—7 responses

Good—18 responses

Fair—7 responses.

In response to the question requesting a description of the major strengths or benefits of the workshops, there were three areas of agreement. First, the participants felt they were challenged to recognize their limited knowledge of the subjects presented. In regard to this point many teachers and administrators commented on the value of having updated information which was based on research in the field. A second point of agreement was that teachers felt that regardless of grade level they had all been given ideas on how to improve the construction of their own tests. Third, an amazing number of the participants commented on the value of the summary

notes which were distributed a week following each workshop. Many commented on the value of the *Teacher's Knowledge Test*¹ which was given to those who desired to take it.

Future Plans

Guidelines for further workshop planning evolved from the responses to the question, "What areas would you like to have emphasized if teacher-made testing workshops should be planned for 1962-63?" These were the most pertinent responses:

1. Work on actual test construction by subject matter areas. (How to write items for good objective and essay questions for our curriculum.)

2. More work on elementary school testing in Social Studies, Science, and Language Arts.

3. Further work on grading and evaluation of tests constructed by teacher.

4. Work on how to construct and use local school norms.

During the summer of 1962 plans were made with E.T.S. to continue the teacher-made testing workshops. A proposal for working with teachers of grades 4, 5, and 6 in the area of social studies was initially developed by the Director of Elementary Education and seven Elementary School Principals. In consultation with E.T.S., plans developed to have all teachers of upper grades work in groups by schools to develop test specifications and to begin to write test items prior to the first meeting with the consultant from E.T.S. Some work will be done by correspondence with specific consultants from E.T.S. assigned to the project. Six meetings have been planned

¹"Teacher's Knowledge Test," Form U-R—TKT-4, University of Illinois Bureau of Educational Research, Urbana, Illinois.

for elementary teachers, four of which will be work sessions in small groups; two meetings will be large-group sessions with the consultants present. The Director of Elementary Education will coordinate the elementary workshops. We are planning to have key school personnel who were involved during the first year's in-service program assume leadership roles in working with teachers to develop the skills of writing test specifications and valid content items. This feature is common to the elementary and secondary programs, even though the two levels will work separately during this second phase of our in-service work. Another common feature, however, will be the use of the summary notes of the lectures presented by consultants during the first year's work, and the distribution and use of two pertinent booklets published by the Educational Testing Service, namely: *Making the Classroom Test—A Guide for Teachers*, and *Short-cut Statistics for Teacher-Made Tests*.

Secondary teachers in the fields of English, Homemaking, Mathematics, Social Sciences, and Science will meet periodically throughout the year by departments to develop test specifications, to write test items, and to meet with consultants. Each discipline will have two meetings with consultants with provision for correspondence between visits. The Coordinator of Secondary Education has the responsibility of facilitating this liaison activity, and for working with secondary principals, curriculum committee chairmen, and department heads who will assume the leadership roles during the work sessions with teachers in the respective secondary schools.

The nature and tone of the second year's work will obviously permit tremendous latitude for teachers to become involved in the selection of projects

which they feel will be of most interest and benefit to them. For example, it may be that English teachers will choose to focus on a reader reliability study and science teachers on the development of a final examination to be administered at the end of the year. Whatever the project, it will be the aim of the administration to help teachers to apply new knowledge and skill to their daily work.

Signals of Need

Do teachers in your school system need to improve their skills in test construction and extend their knowledge about testing? There are signals to help you to decide whether teachers need or want help in developing greater competency in the measurement of educational attainment.

1. Teacher relies too much on own standards of judging educational attainment; does not check judgment against opinion of other teachers in the same subject area.
2. Teacher's tests place an overemphasis on details to the neglect of basic principles.
3. Teacher tends to put off test preparation and does not inform students early enough as to kinds to expect.
4. Teacher contributes to or actually creates anxieties related to test administration.
5. Teacher gives tests which are too short and inefficient in form.
6. Teacher is irresponsible or indifferent to setting the stage for testing and fails to provide adequate security for tests.

It is apparent that it is the obligation of every supervisor, curriculum worker and administrator to be aware that effective education requires effective measurement of educational attainments. We need to be increasingly concerned with behavior that can be measured, and at the same time discerning enough to recognize the abstractions which are not im-

(Continued on page 72)

Testing

(Continued from page 10)

and studied their language behavior as manifested in imaginative stories. The more creative children wrote longer stories; used a greater number of different words; more frequently used first person pronouns, verbs like "said" and "was," and conjunctions indicating cause or consequence; and the like.

In conclusion, we recognize that the concepts of creative talent which have guided the work described in this article will not satisfy many individuals. Nevertheless, it seems rather clear that high scores based on the measures described identify individuals who behave in ways commonly regarded as creative and that individuals who behave in ways commonly considered as creative achieve higher scores than similar individuals identified as behaving in relatively uncreative ways. Although there is a need for much developmental work, present instruments can be used productively in experimental and research projects.

References

Elyse S. Fleming and S. Weintraub. "Attitudinal Rigidity as a Measure of Creativity in Gifted Children." *Journal of Educational Psychology* 53: 81-85; 1962.

J. W. Getzels and P. W. Jackson, *Creativity and Intelligence*. New York: John Wiley, 1962.

J. P. Guilford and P. R. Merrifield. *The Structure and Intellect Model: Its Uses and Implications*. Los Angeles: University of Southern California, 1960.

C. W. Taylor. "The Creative Individual: A New Portrait of Giftedness." *Educational Leadership* 18: 7-12; 1960.

E. P. Torrance, *Guiding Creative Talent*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1962.

H. R. Wallace. "Creative Thinking: A

Factor in Sales Productivity." *Vocational Guidance Quarterly*, Summer 1961; p. 223-26.

P. S. Weisberg and Kayla J. Springer. "Environmental Factors in Creative Function." *Archives in General Psychiatry* 5: 554-64; 1961.

All Teachers

(Continued from page 29)

mediately measurable. We must place value on the statements of measurement specialists who tell us that the key to improvement of classroom tests is objective analysis of test content and performance in relation to a set of explicit standards for test quality set by the test maker.

All teachers have to make tests and all can be helped to improve—but not without extra effort, extra time, and sufficient money to secure the services of test experts. Critical to the problem is finding the most promising career teachers to involve in the program; teachers who are themselves interested and influential in communication with other teachers. Further, I would argue strongly that an in-service training program on testing must be carried on over periods longer than a year—perhaps two or three. The payoff could only be better teaching.

References

Paul Diederich. "The Problem of Grading Essays." Princeton, New Jersey: Educational Testing Service.

Robert L. Ebel. "The Nature of Modern Tests." Princeton, New Jersey: Educational Testing Service.

"Making the Classroom Test." Educational Testing Service. Evaluation and Advisory Services, Series No. 4. Princeton, New Jersey.

Dorothy A. Wood. *Test Construction*. Columbus, Ohio: Charles E. Merrill Books, Inc., 1961.

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