

Shall Tests Go Out of the Window?—

GUY M. WILSON

No, tests should not be discarded altogether, our author assures us, but he does feel that there is room for growth of concept. The editors believe that much rethinking of this process is necessary and we cannot solve the problem by failing to meet it. Therefore, we present this article by Guy M. Wilson, Professor Emeritus, School of Education, General College, Boston University, Mass., which suggests a point of view of the function of tests and testing in the modern school.

IT IS THE TEACHER who is the key individual in the actual development of any school program. She is the one who comes into immediate contact with the pupils, for whom the school exists. Parents represent a wealth of concern and interest on the side of the pupil. Administrative and supervisory officers are on the side of the teacher, supposedly to make her work more pleasant and effective. But it is the teacher who is the key to what happens to children. Any research or administrative programs which reach the classroom and the child should be arranged cooperatively with the teacher.

As the teacher rises in responsibility and power, her obligations increase. She is no longer a convenient servant, covering a predetermined course of study by specified methods, or in extreme cases, covering specified pages in specified textbooks. Books should be at hand—texts, supplementary books, reference books, free reading books. Means for developing firsthand experiences should be available. Curriculum bulletins cooperatively produced by teachers, principals, supervisors, and curriculum specialists should furnish guides, but always they are merely suggestive. All materials must be constantly supple-

mented by new references, new illustrative units, and reports on superior work of teachers. The needs of the children, as interpreted by their teachers, and by the pupils themselves, are the important factors in determining what shall be done from day to day in particular schoolrooms.

Tests enter into this picture as service tools, potentially useful as aids in discovering the needs (and accomplishments) of the children. Can tests be as fluid, as spontaneous, as adaptable as the needs of the children require? They cannot without much help from the classroom teacher. She must change and adapt, or create anew, as she wisely interprets the needs of her pupils. This means that most tests used will be informal and teacher-made.

It can be predicted that at least two-thirds of school work does not lend itself to testing by the use of standardized tests. Since standardized tests to date, notwithstanding efforts to the contrary, emphasize drill material and the drill viewpoint, such tests are unacceptable in many phases of the school program. Any tests dealing with problem situations or children's appreciations must be informal and teacher-made. For much of this work *no tests are needed*.

Teachers Express Their Views

The view that the teacher is of major importance in the testing program is readily understood and accepted by teachers. One teacher says:

Tests are of little value if given merely for the sake of having a testing program. Such tests are usually given near the close of the year when they can serve little purpose except general comparisons.

Useful tests are fairly frequent and reasonably short in length. They should be used to show the child where he needs to place more emphasis. The record should be so kept that the child may see his improvement. Tests should be planned as a benefit to the child's progress, and so used and kept that he may see their value to himself.

Another writes:

Teachers are interested in pre-tests that reveal the phases of the work to which pupils need to give attention. They should help the teacher in planning for individual differences. Tests can help in revealing to pupils and teacher the next steps in learning and teaching. General batteries of tests given at the close of the year have little value for anyone. Frequently they do not give good general appraisals. They tend to transform problem and appreciation subjects into drill subjects, if the questions are taken seriously by the teacher.

Still a third expresses this opinion:

Truly functional tests need not be kept secret. They may be open to inspection by the pupils at any time. In drill subjects, they will tend to provide complete coverage and hence will fully indicate the drill load. In problem situations, they should give free choice among many large problem opportunities and in no case call for minor details. Hence, there is every advantage in having children know the examination long in advance. In appreciation subjects, only personal likes and interpretations should be called for, and there should

be no surprises; pupils and teacher together might properly determine any examination well in advance of the time for writing, if writing is called for.

Determiners for Testing

Obviously, tests to be functional, truly useful in the schoolroom, must be constructed with great care and discrimination. Among the important factors that should operate to determine and change the form and content of tests are the following, listed somewhat in the order of their value:

The pupil and his needs must, by all odds, be listed as the first and great factor. In a second grade spelling list of a standard test, I find the words door, seed, winter, gate, broke, wagon. Are these good words for a particular second grade? The answer must come from the children. Are the words known by the children of my room well enough to come easily into spontaneous writing? Have the children of my room studied all of these words for mastery?

For almost any second grade, the answer to both these key questions would be in the negative. Therefore the teacher should go through the list with a blue pencil, freely marking out the inappropriate words. There is no advantage in confronting children with end-tasks on which they are sure to register a high degree of failure. *Failure is not educative; it is not good mental hygiene; it is not good teaching.*

The kind of learning experience which the teacher has provided for children is the second factor for determining the content and form of a test. What has been done; what has been taught? How has it been done; how has it been taught? If the child is not to be

confused, dazed, and stunned, these are legitimate questions.

In a primary battery of a standard test, second grade children are expected to read understandingly (and be tested for a grade, possibly for promotion) about queen bees, workers, and drones; about telephones, electricity and electric irons; about weapons, engines, and other machines in a museum. Whether or not this is appropriate testing depends upon what the teacher has done, what the community affords in the way of experience, and what the children have done.

These points are probably clearer in an arithmetic test for second grade children in the same battery. In the fundamentals, the children are expected to add to five addends with dollars and cents, to subtract to four places, including zeroes and double borrowing, to do one-place multiplying into zero, to do simple short division, and to take one-half of eighteen. This represents a considerable amount of drill, none of which is good for a typical second grade; informal, informational units would be better.¹

In the problems section of the arithmetic test in this battery, children are given opportunity to use judgment (if the work is properly done) in buying a pony, in buying a calf, in buying milk and bread, in buying hay, in buying a camera. These tasks are quite unreasonable for second grade children in general, and unless the teacher has taken up this particular work with these children, the items should be deleted from the test when and if given.

The third factor for determining the form and content of a test is the goal toward which learning in a particular subject or field is planned. This point has been emphasized again and again throughout this article. But it cannot be said too often that a wrong testing program can defeat legitimate aims. It has succeeded in doing so in nearly every subject—in spelling, in arithmetic, certainly in history, geography, science, and literature. *The power to test is the power to destroy.* Standardized tests may frequently be made to aid in teaching the drill subjects, although wrong tests can be detrimental here. But in problem situations and in developing appreciations, standardized tests are generally detrimental and probably will continue to be detrimental; tests, if any, in these subjects must be informal and teacher-made. Unless a test reinforces the final goal, it is not a good test.

The method used as most appropriate for learning is the fourth factor for determining the testing program in a subject. The points made above are applicable here. To date, in general, the whole influence of standardized tests is to emphasize drill. As noted above, in at least two-thirds of school work this is the wrong method. Appreciation experience should frequently go without any testing, except the informal testing of class discussions. Ability in problem solving can not be tested legitimately by drill-type tests. Unless the requirements of right teaching methods are observed, the test should be discarded.

Should statistics or standardization be listed as an important factor in the teacher's testing program? Probably not, or if so, at least it is a subordinate factor, the fifth in the list, with four factors tre-

¹ Guy M. Wilson, "New Standards in Arithmetic", *Journal of Educational Research*, 22: 351-360, December 1930.

mentously important standing above it. Statistics is a tool; it may sometimes be used for polishing and refining; it may aid in interpretation; it should never be used as a primary determiner. Used as if it were the chief determinant, it leads to outstanding blundering. For example, a third grade child tries an arithmetic reasoning test or problem scale; there are twenty problems; he fails on half of them; he gets an equated score of 36; and this gives him a grade equivalent of 3.8.

This statistically determined testing program and summary is wrong from start to finish, and its general influence is bad. In the first place, it misses the purpose of problem work in arithmetic; it reduces valuable functional problem opportunities to words and cues and drill, all of which are wrong. In the second place, it presents chiefly an opportunity for confusion and failure. In the third place, the determination of a 3.8 grade placement misses the point of true instruction and also misses the spirit of the modern school, in which failure is replaced by success.

A Plea for Sanity

The above discussion on tests has sought to loosen the strangle hold of the standardized testing program, to aid in the movement for good mental hygiene and humane treatment of children in all teaching, and to encourage the teacher who takes her teaching task as a sacred trust in terms of child needs today and adult and national needs tomorrow.

There is the constant tendency in our profession to exploit one idea at a time, and to hail a new idea as if it were a panacea. Our profession would look saner to others, and would actually

show greater sanity, if we kept the historical viewpoint on schoolroom practices, recognizing the good in the old, and merely adding anything of value which the new might develop. Many teachers and supervisors condemn drill because it is old. It was applied too widely, too ruthlessly; but it has high value in limited areas—when preceded by meaning, motivation, and pupil-teacher planning.

Some teachers and supervisors now appear to be at the point of throwing overboard all types of testing. There is no doubt that standardized testing in its panacea days was indiscriminating and did much damage, particularly to pupil and teacher initiative, and in the fields where problem and appreciation techniques are properly applied. But surely any teacher can acquire viewpoints and values from a study of the testing movement, and she should be more resourceful in evaluation because of such study. A discussion of "evaluation" today which omits all reference to testing, standardized, new-type, formal, informal, or essay—is showing the shortsighted policy which historically has characterized our profession. We want to evaluate adequately and with full pupil cooperation and understanding. This can be aided by "observation, anecdotal records, reports to parents, group planning." But these newer procedures in the hands of the average teacher are not always adequate. They can be aided by judicious use of other procedures, some of which date back a few years. New terms represent partly a natural growth, and partly a desire for new terms in order to escape the misunderstandings attached to older terms. But the well-read teacher should

know the essential unity of many terms.

Gradually we are realizing that the traditional curriculum isn't a good guide, nor a good general framework. Life is much broader than the traditional curriculum. A functional viewpoint on education must take the child where he is and advance him in

line with his background, experience, and ambitions; and his morale must be maintained, his personality preserved and further developed. The experience viewpoint is the only profitable one for the individual child. He is the chief determiner of good evaluation procedures.

Workshop Designing

W. GEORGE HAYWARD

Continuous appraisal is necessary in planning experiences for adults as well as for children. W. George Hayward, principal, Elmwood School, East Orange, N. J., and director of the Lehigh University workshop describes how constant evaluation leads to improved learning for adults.¹



Courtesy San Francisco (Calif.) Public Schools

Group Planning for Good Growth

THE LEHIGH UNIVERSITY Summer Workshop for 1947 is being planned. It is the third successive season in which such planning has moved forward. The planning which is now in process has been changed each year but has continually supported two basic policies. First, the program should be planned to meet the needs of the teach-

ers who will participate and second, the program and requirements should operate in such a way as to encourage carry-over throughout the following school year to the highest degree possible.

Planning With Those We Serve

In planning the 1945 workshop, which was a cooperative effort between the university and the Bethlehem school system, numerous conferences were held. Preliminary conferences, in which

¹ In view of forthcoming plans for summer workshops, we feel that this account is timely. We shall publish three more articles on the same subject in next month's issue.

Copyright © 1947 by the Association for Supervision and Curriculum Development. All rights reserved.