

Test Construction and Evaluation of Student's Achievement

B.T. Lahane

Principal, Sambhajirao Kendre Mahavidyalaya, Jalkot Dist.Latur, MS, India

Abstract

The achievement can be measured in the field of behavioral sciences as well as humanities. It means that the values in professional education know only what their measuring instruments reveal when they seek to determine the individual differences present in learners and the degree to which the behavioral patterns of these learners change as a result of educational experiences. What matters in the evaluation is reliability and validity. Educational objectives are prominently measured. The teacher should have clear concept of evaluation. He needs many tools to assist him in this process. These tools are frequently tests, but they can also be non test evaluation. The teacher must provide an opportunity for the student to display his behaviour with regard to the goals under consideration. In any event, the student behavioral patterns must be displayed, and the process of eliciting these behavioral patterns may be such that they are obtained in a natural situation or in an artificial situation. There is considerable similarity between the procedure used to construct standardized achievement tests and that used to construct standardized aptitude tests.

KEYWORDS: behavaour, pattern, reliability, validity, domain, test, aptitude, evaluation

Introduction:

Many years ago Lord Kelvin, the British mathematician and physicist, was reported to have said, "When you can measure what you are speaking about and express it in numbers, you know something about it." When Lord Kelvin made this statement, he probably was not addressing it to the area of psychology; yet, what he said with respect to mathematics and physics can also be said of the behavioral sciences. Moreover, even though he made the statement at the turn of the twentieth century, it is as true now as it was then.

What does Lord Kelvin's statement in terms of psychological testing and its impact on the schools and universities of today? At the very least, it means that those interested in professional education know only what their measuring instruments reveal when they seek to determine the individual differences present in students and the degree to which the behavioral patterns of these students change as a result of educational experiences. In other words, the degree to which we know and understand the educational process and its impact upon the student is determined to a large extent by the degree of reliability and validity of the psychological tests used in the educational process, that is, the degree of consistency of the test scores and the degree to which they serve the purposes intended.

It is well-known that the degree of reliability and validity of the psychological tests commonly used in education is much less than that typically present in the measuring instruments used in the physical and biological sciences. However, in many instances there exists a sufficient degree of reliability and validity to allow evaluations to be made of pupil changes in behavior and, ultimately, of the effectiveness of educational programs.

The evaluation process is completely tied to educational objectivities. The teacher must necessarily teach and evaluate in terms of his educational objectives. If a teacher has relatively few goals and these are of a relatively straightforward nature, his evaluation problems are simplified. On the other hand, if a teacher has a great variety of goals and some, or all, are complex, his evaluation problem becomes more difficult.

The teacher begins the evaluation process by formulating a clear picture of the desired student behaviour which will be exhibited if the objective has been achieved. He must also have, of course, a clear concept of other degrees of achievement of the objective, insofar as they are reflected in observable pupil behavior. His task now is to identify the present state of the pupil behaviour with regard to the objective under consideration. He needs many tools to assist him in this process. These tools are frequently tests, but they can also be non test evaluation instruments, for example, rating scales and anecdotal records. It is sometimes difficult to decide which tests or which non test instruments will yield the best picture of the current state of the pupil's behaviour. If possible, several measurements or observations will be made. All too frequently, however, the teacher finds himself in a position whereby he has no evaluation instruments which are totally suitable. Hence, he necessarily improvises and generally reduces the degree of certainty with which he interprets the results of the administration of the evaluation instrument.

Having identified certain goals in terms of which he wishes to evaluate and having identified one or more evaluation instruments to use, the teacher must now provide an opportunity for the student to display his behaviour with regard to the goals under consideration. In the case of tests, this is as simple as providing an opportunity for the test to be administered. In the case of non test instruments, providing the opportunity becomes more difficult in many instances. In any event, the student behavioral patterns must be displayed, and the process of eliciting these behavioral patterns may be such that they are obtained in a natural situation or in an artificial situation.

Tests in the cognitive domain with respect to recall of information are probably of similar effectiveness. As we travel through the cognitive domain from the simplest category, knowledge, to the most complex category, evaluation, we find that utility of paper-and-pencil tests has decreased. However, the ingenuity of test builders is such that the loss of utility is far less than many suppose. Both essay tests and some of the more elaborate types of objective tests are able to reveal pupil behavioral patterns in the cognitive domain which are of a complex nature.

Since the cognitive domain is clearly the most important of the three domains, it is not surprising to find that paper-and-pencil testing has dominated the educational scene. To a large degree this is justified. On the other hand, we must remember that non test devices may, in spite of their technical deficiencies, yield high quality information on the basis of which evaluations can be made. This is true primarily because they can be applied within situations which tend to be less artificial than some of those in which the paper-and-pencil tests are used. Certainly, paper-and-pencil tests should be used only when they will provide the most meaningful information to the student behavioral patterns being evaluated.

Paper-and-pencil tests are suitable for providing data on the basis of which a teacher can evaluate the degree to which students have achieved educational objectives of a verbal or mathematical nature. In other words, they are suitable when applied in the cognitive domain. The cognitive domain, is composed of six subdivisions, the first of which is knowledge. Knowledge, in essence, means the

ability to recall information upon request. Determining the degree to which students can recall information has been a traditional task assigned to paper-and-pencil tests. The remaining five subdivisions-comprehension application, analysis, synthesis and evaluation may be thought of as parts of a single category known as understanding. The term 'understanding' suggests that part of the cognitive domain which is represented by objectives other than those requiring the recall of information per se.

There is considerable similarity between the procedure used to construct standardized achievement tests and that used to construct standardized aptitude tests. Hence, it is appropriate to consider these procedures simultaneously. They can be summarized within three main categories: planning and writing the test, pretesting and analyzing the results, and forming the test.

Standardized tests are the outer growth of the intense and widespread testing need on the part of educational institutions with respect to a particular problem or set of problems they are facing, for example, as in the case of determining changes in a student's command of his basic skills as he progresses through elementary school. Only when the need is of this nature is there sufficient justification for the expense and effort of constructing a standardized test. As the need crystallizes, it, in effect, forms the program or outline for the test.

The situation is no different in formal education. It is necessary for teachers and school administrators to make judgments concerning student potential, particularly academic potential, and to organize teaching programs and to evaluate student success on the basis of these appraisals. Fortunately, as in the case of business and industry, it is not necessary to rely on informal techniques in order to obtain needed information concerning potential. A large variety of reasonably well-developed and standardized instruments are available which are capable of yielding data on the basis of which evaluations can be made of human potential in a wide variety of areas.

In education the scholastic aptitude of a student is clearly the most vital aspect of his total potential. However, one cannot ignore other aptitudes such as mechanical, musical, artistic, clerical and spatial. Tests are available in all of these areas, but only in the case of scholastic aptitude are they widely administered. The less important areas are of interest to teachers and counselors in the case of only a part of the student body with which they are concerned. Hence, administration of tests in these areas in a wholesale manner is relatively uncommon. Ordinarily, such tests are administered to students only in the case of special need.

An aptitude is an individual's potential for learning a given skill when he is provided with appropriate instruction. The aptitude family of student characteristics is certainly different from his achievements, which are his current status with respect to the skill in question. He can have relatively low achievement and, simultaneously, relatively high aptitude in a given respect.

In terms of aptitude tests and achievement tests, the differentiation is by no means clear. The content of tests in these two areas sometimes is so similar that one is unable to differentiate between aptitude and achievement tests items when lifted out context. This problem stems from the fact that test builders commonly assume that the best way to determine one's potential for a given future skill is to examine his achievement in that area of related areas today. Consequently, some aptitude tests, particularly scholastic aptitude tests, strongly resemble achievement tests. Indeed, the tendency for this to occur seems to become stronger as the years pass.

Scholastic aptitude tests (or general mental ability tests, as they are sometimes called) cover a wide variety of content. Cronbach describes the variability of the test

content of general mental ability tests by identifying the degree to which the tests are measures of the outcomes of education. He classifies the degree of educational loading into six categories: (1) Subject-matter proficiency; (2) General educational development; (3) Reading, vocabulary, arithmetic reasoning; (4) Verbal analogies, number series; (5) Figure series, matrices; and (6) Learning samples. The maximum educational loading is, of course, the first category, subject-matter proficiency, whereas the minimum educational loading is the last category, learning samples. Tests which emphasize the first category are ordinary achievement tests devoted to a specific body of subject matter. Tests which emphasize the last two categories deal with tasks which involve concepts and experiences familiar to all examinees but which still challenge the reasoning power of that person. In the case of these tests, the emphasis upon language and culture is relatively small.

Many of the mental ability tests used in formal education today cover more than one of the six categories. Usually, they do not emphasize the last two categories. Instead, they concentrate on the middle of the list, that is, on the reading, vocabulary, an arithmetic reasoning category, as well as the verbal analogies and number series category. Several important tests overlap the second category, the general educational development category, and thereby create confusion when one attempts to differentiate between achievement and aptitude tests in terms of test content. In this respect the overlap between an achievement test and a corresponding mental ability test can be very great.

Conclusion:

In summary, the achievement tests and aptitude test used in formal education are identified primarily on the basis of their purpose, rather than on the basis of their test content ability tend to overlap a great deal is a less serious matter than often supposed. If the tests in the respective areas serve, to a satisfactory degree, the purposes for which their use is intended, then no serious doubts should exist with regard to their utility.

Works Cited

1. Ahmann Stanley,(1981) 'Testing Student Achievements and Aptitudes' Prentice Hall, New Delhi
2. Wood Dorothy A, (1989)'Test Construction', Ohio Press
3. Bloom Benjamin S, (1978)' Taxonomy of Educational Objectives', Longman New York
4. Ebel Robert ,(2001) 'Tests and Examination', The Macmillan, New York.
5. Harris David P,(2000) 'Constructing the Test...' Withrop Publishers.