Wright, B.A.. **Physical Disability. A Psychosocial Approach.** New York: Harper and Row Publishers, 1985.

WHO. **Training the Disabled in the Community: Guide for Policymakers and Planners.** New York: WHO, 1980.

---------- // ----------

FACTS AND FIGURES

Percentage of Female Participation in Higher Education by Program Type and Level (1989/90)

	Program					
	Regular		Evening		Kirest	
Participation	Dip.	Deg.	Dip.	Deg.	Dip.	Deg.
Enrolment	15.0	8.1	33.3	22.6	6.5	14.1
Graduates	13.3	8.4	37.9	27.4	10.0	10.3

Source: Higher Education Main Department, **Facts and Figures**. Jan. 1991 (p. 21)

*IER Observes: As shown in the Table, a greater percentage of females joined extension programs rather than the regular diploma and degree programs. It would be interesting to know why this is so.

---------- // ----------

Wright, B.A.. **Physical Disability. A Psychosocial Approach.** New York: Harper and Row Publishers, 1985.

WHO. **Training the Disabled in the Community: Guide for Policymakers and Planners.** New York: WHO, 1980.

FACTS AND FIGURES

Percentage of Female Participation in Higher Education by Program Type and Level (1989/90)

	Program					
	Regular		Evening		Kiremt	
Participation	Dip.	Deg.	Dip.	Deg.	Dip.	Deg.
Enrolment	15.0	8.1	33.3	22.6	6.5	14.1
Graduates	13.3	8.4	37.9	27.4	10.0	10.3

Source: Higher Education Main Department, **Facts and Figures**. Jan. 1991 (p. 21)

*IER Observes: As shown in the Table, a greater percentage of females joined extension programs rather than the regular diploma and degree programs. It would be interesting to know why this is so.

A TENTATIVE GUIDELINE TO THE CONSTRUCTION OF A GOOD CLASSROOM TEST (PART ONE)

ABRAHAM HUSSAIN

I. Introduction

According to some studies (e.g. Nelson, 1939) on the history of measurement and evaluation, it was probably during the time when people realized the advantage of expressing the notions of length or of distance that they began to make comparisons between the length to be described and things with which everybody was familiar. Various parts of the body such as the forearm were commonly used for this purpose. Authorities in the field assert that in practice this "unit" of length was defined as the bent forearm from the elbow point to the finger tip, in which case length was measured in so many "forearms". Of course, as a unit of measurement the forearm had its own shortcomings, because not all forearms were of the same length.

Obviously, a similar problem is faced in the teaching-learning process How can we measure pup comprehension of cert material? How do construct tests? What do want a particular test measure? To answer these similar questions, one ca use yards, pounds or meters, let alone forearms; evaluation in education is more complex, yet it is an indispensable tool for the teacher.

It is widely believed that measurement and evaluation in education can involve both quantitative and qualitative judgment.

A teacher of English, for instance, might ask his pupils to write a composition. In doing this, he perhaps has in mind two general qualities to be evaluated - the mechanical features of the language in

the composition and the content of the subject-matter. In evaluating the mechanical part, the teacher might think of such things as spelling, punctuation, capitalization, paragraphing and coherence. When he has finished reading the paper, he may record the number of errors of a mechanical nature and then decide whether on the basis of such errors, the composition should be marked "80" or "90", or some other value. When he thinks of the content, the teacher might have in mind either the other composition papers written by the same class or some general notion which he has gathered from proceeding experience as to what the content of a composition on this particular subject ought to be; then he may tell the quality of the composition according to his best judgment.

In this procedure of evaluation, Nelson, (1939) believes that one might note:

(a) the teacher has no definite "units" in the "scale" which he uses in evaluating a composition;

(b) he may not use a scale at all except a rather vague general notion which he has in (his) mind;

(c) he may consider certain objective features of the composition such as spelling and punctuation, which may be called "units" in his "scale";

(d) the scale which he uses might have been built up out of his own experiences in evaluating composition;

(e) the teacher may not intend to measure all of the pupil's ability in composition but may give attention to one sample.

One way or another, some form of evaluation is indispensable in teaching. The purpose of a test, methods of constructing and administering a test, and ways of correcting test papers are of paramount importance in teaching-

learning processes. Evaluation requires knowledge, skill and talent. Depending upon a particular goal, one could think of several types of tests such as aptitude tests, personality tests, interest inventory tests and so on. However, because the primary objective of this paper is to highlight (for educators and school teachers) the importance of a "good test" in the ultimate improvement of teaching and learning, it focuses specifically on constructing **classroom tests**.

Evaluation in this context could be defined as the systematic process of determining the extent to which educational objectives are achieved. This definition generally assumes that evaluation is a systematic process which underlies the planned and controlled observation of pupils, and that educational objectives (goals), without which it is practically impossible to judge the extent of pupils' progress, have been previously identified (Gronlund, 1976).

Classroom tests are constructed in the light of certain general principles.

II. General principles of classroom tests construction

In constructing classroom tests, it is quite possible that by focusing our attention narrowly on the detailed procedures of constructing specific types of test items, we may lose sight of our major purpose, which is to develop a valid instrument for evaluating pupils' achievements. As pointed out by some authorities, this tendency could be avoided by careful preliminary planning and by the use of a general frame of reference within which one views the specific procedures of test construction.

According to Gronlund (1976), the following principles of classroom testing provide such a frame of reference:

1. Test construction procedures should take into account the purpose of the test. If we are interested in determining a pupil's readiness to start a new unit of work or a new course, for instance, our test would generally be confined to a limited area of knowledge or skill. If on the other hand, our test results are meant to evaluate a pupil's progress towards identified educational objectives, one should naturally be interested in constructing a general achievement test. Such a test usually enables us to rank pupils according to their achievement and to identify their general areas of weakness. For such purposes, one would want a test:

(a) that measures a representative sample of the course objectives and course content,

(b) that is difficult enough to provide a reliable ranking of pupils, and

(c) that contributes to improved teaching and learning.

2. The types of test items used should be determined by the specific learning outcomes to be measured.

In constructing a classroom test, one of our major concerns is that the test items call forth the particular abilities (e.g. defining terms, stating facts, applying knowledge, identifying cause-effect relations) indicated in the specific learning outcomes pertinent to each of our course objectives. This is said to be particularly necessary if one is to accept a pupil's responses to the test items as evidence that the specific learning outcomes, and consequently the course objectives, have been achieved.

Whether a test item actually measures the particular ability called for

in a specific learning outcome depends to a large extent, of course, on the skill with which the test item is constructed.

3. Test items should be based on a representative sample of the course content and the specific learning outcomes to be measured.

No matter how extensive a test may be, only a limited sample of the many possible test items can normally be included. Suppose we expect pupils to know thousands of specific facts, yet we test only a limited number of them; we expect pupils to develop understanding applicable to innumerable situations but we can test its application in only a limited number of situations; and we expect pupils to develop thinking skills which will enable them to solve a variety of problems, but again we can test their problem solving ability with only a limited number of problems. In each area of content and for each specific learning outcome, then, we merely select a sample of pupil behaviour and accept it as evidence of achievement in that area. We assume that the pupils' responses to our selected set of test items are typical of what that response would be with other test items drawn from the same area. This, of course, means that our limited samples must be selected in such a way as to provide a representative sample in each of the various areas for which the test is being developed.

4. Test items should be constructed in such a way that extraneous factors do not prevent the pupil from responding correctly.

When a pupil has achieved a particular learning outcome, we naturally expect him to obtain correct answers to those test items which measure its attainment. Both the teacher and the pupil would be unhappy if he or she answered such test items

incorrectly merely because the sentence structure was too complex, the vocabulary too difficult, or the type of response called for too vague. Such factors, which are extraneous to the central purpose of the measurement, limit and modify the pupil's responses, and prevent him from showing the true level he has attained.

One way of preventing extraneous factors from distorting our test results is avoiding ambiguity. Objective test items are especially subject to misinterpretation where long complex sentences are used, where the vocabulary is unnecessarily difficult, and [illegible] words which lack [illegible]se meaning are used. [illegible] the remedy for this [illegible] to be careful in the [illegible]e of words, from the point of both the [illegible]l of reading difficulty [illegible] preciseness in meaning, and in the use of brief and concise sentences.

5. Test items should be so constructed that the pupil obtains the correct answer only if he has attained the desired learning outcome. Here, one is concerned with those factors which make it possible for the pupil to respond correctly even if he lacks the necessary achievement. We are talking here about clues, some rather obvious and some very subtle, which unintentionally creep into test items during construction. These can lead the non-achiever to the correct answer and thereby prevent the items from functioning as required.

The most obvious clues in test items as indicated by Gronlund (1976) are probably those which are linked to grammatical structure. In the following example, for instance, notice how the article "an" could provide a direct clue to the answer.

- A porpoise is an
  A. Plant C. animal
  B. reptile D. bird

- A piece of land that is surrounded by water is know as an_____.

For this item, the two most plausible answers are "island" and "peninsula". Since peninsula begins with a consonant and does not follow the article "an", it is ruled out as a possibility. This, of course, does not imply that pupils need to know the rules for good grammatical structure, in order to use such clues. In fact, most clues are not analyzed and evaluated as above. Rather they are responded to in terms of partial knowledge and hunches. "An peninsula" just does not sound right to the pupil, so he responds with the word "island" and obtains the correct answer.

Leads to the correct answer may also be provided by simple verbal associations. The word "wind" in the following example provides a clue to the answer.

Which one of the following instruments is used to determine the direction of the wind?

A. Anemometer C. Hygrometer
B. Barometer D. Wind vane

Rather than leading the uninformed to the correct answer, some clues can also lead the non-achiever away from the correct answer. In the following case, for instance, the same clue could make "wind vane" a plausible (but an incorrect) answer for those pupils who have not learned the uses of the various weather instruments.

Which one of the following instruments is used to determine the speed of the wind?

A. Anemometer C. Hygrometer
B. Barometer D. Wind vane

III. Types of classroom tests

Although a wide variety of tests exist, there are, in general, only two major categories of tests-Objective and Subjective (or essay).

Objective tests

In general, objective tests are believed to reduce the opportunity for subjective interpretation by the person taking them.

Objective test items can be classified as:

A. True-False requiring a single answer;

B. Matching and

C. Completion, which also allow a limited choice of responses.

All of these objective tests have the same general characteristics:

(a) the pupil operates within a completely structured situation;

(b) selects an answer from a limited number of choices;

(c) responds to each of a large smple of items; and

(d) receives a score for eache answer according to a predetermined key.

True-False test

Critics say these are widely used because they ar quick and easy to implement and that teachers tend to use them "to take the easy way out". However, if constructed with care, True-False items can be used to assess higher cognitive functioning, and they also have the advantage of sampling a large amount of subject matter. Since pupils can answer two or three times as many true-false items as other types in the same amount of time, true-false tests give pupils a greater opportunity to indicate what they have learned.

Good true-false test items should be written according to certain specific principles:

(1) **Keep them simple**

Each true-false item should be stated in the simplest, most direct terms, which reduce the possibility of "reading into" the item more than one meaning or interpretation. It is important that teachers should know precisely what stimulus they wish to present and which response they wish to obtain. For example, if pupils are not sure of what they are being asked to do because an item is ambiguous, they will not have a clear-cut opportunity to demonstrate that they know the answer. In this case the item will have lost its objectivity.

Let us take a look at the following illustration by Bertrand and Cebula (1980) which assumes that a unit on electricity has been covered by the teacher.

T.F. A fuse is placed in our home because we want to be safe from the dangers of electricity.

What dangers? Where is it placed in the home? The item is not specific enough to the subject matter and offers the pupils opportunities to think about extraneous matters. A better way of writing the item would have been:

T.F. Fuses prevent too much electricity from passing through the wires.

Although the improved version of the item does not include the dangers of overloading, the next item does.

T.F. Too much electricity in the wires will cause them to get very hot.

From here the test could include a series of simple statements, each referring to a specific bit of knowledge covered in the unit and each presenting a single clue to which the pupils can respond.

(2) **Avoid words that can be interpreted in different ways.** Whenever words are included that are ambiguous in degree or amount, objectivity could be lost and pupils will produce different possible answers. For instance:

T.F. It does not take very long for a wire to become over-heated.

The phrase "very long" is open to too many interpretations and is thus misleading and confusing.

(3) **Do not include clues to the answers.** Words such as "never" or "always" usually signal that the answer is false, while words like "sometimes" or "normally" often indicate a true statement. For instance:

T.F. Too much electricity in the wires **usually** causes them to over-heat;
or
T.F. A fuse **always** prevents too much electricity from passing through the wires.

Writing good true-false items should not be done a few minutes before administering the test. Good true-false items require time for reviewing instructional objectives and for materials that have been taught. It is only then that teachers can translate the subject matter into a test that will effectively measure pupils' learning behaviour.

Multiple - Choice test

The Multiple - Choice item is felt to be more difficult to construct than the other objective tests. However, a look at its basic elements shows that this is not really so as long as teachers pay attention to a few simple rules. Each multiple-choice item consists of a stem and a series of alternative responses, one of which is the correct response.

The stem: Like other objective test items, here also clarity is the first and most important consideration. In a multiple-choice question, the stem should present the problem so clearly that pupils will know precisely what is expected of them. It should be constructed in such a way that it leads directly to the alternatives without ambiguity. This can be assured if both the stem and the alternatives are written as grammatically complete statements. For instance:

The Blue Nile originates from Lake Tana in North-western Ethiopia.

Stated this way, the entire item can easily be given a clearly stated stem and a good set of alternatives. For example the sentence can be broken down in the following way to construct the alternative responses or "distructors":

The Blue Nile originates form Lake Tana in:

A. Northern Ethiopia
B. Western Ethiopia
C. North-western Ethiopia
D. North-eastern Ethiopia

It does not really matter very much where the stem is actually split, so long as it makes good sense and contains most of the information. Items at this level should provide clues for accurate recall in order for the pupils to be accurate in their selection of the answer. It does not matter, either, whether the stem is written as an incomplete sentence, as above, or whether it is restated as a question. For instance, after the initial statement about the Blue Nile River, the multiple-choice item could also be written as follows:

Where does the Blue Nile River originate?

A. In Northern Ethiopia
B. In Western Ethiopia
C. In North-western Ethiopia
D. In North-eastern Ethiopia

The alternative The alternatives (sometimes called options) are the "multiple-choices" among which pupils select. Each alternative must be written to flow logically and grammatically from the stem, and each should appear plausible; that is, a pupil who does not recall the exact origin of the Blue Nile River (in this case), might choose any one of the alternatives. Thus, multiple-choice items with plausible alternatives demand more intensive "remembering" to avoid guessing. The "distructors", although plausible, must have no element of correctness about them. They must be clearly and unequivocally wrong responses so that pupils will not be misled by choices in which they can see partially correct elements. This can make the question ambiguous for the pupil and unreliable for the teacher as a means of evaluating learning. Moreover, experts on test construction stress that keeping the responses as short as possible also helps to avoid ambiguity. Finally, all choices should be homogeneous in content to avoid any that are obviously wrong.

Without considering possible factual error, here is an example of an item that has been poorly written:

The rich alluvial soil of the Blue Nile Valley could produce one of the most important regions of forestry in the nation, which will be formed by:

A. the creation of soil by the erosion of the surrounding mountains
B. the depositing, by the river, of silt along its banks
C. annual flooding
D. the construction of dams and waterfalls

The item-stem is too long and is poorly written. Pupils do not know whether they are being asked to say how the alluvial soil was formed, or how the Blue Nile Valley was formed. Although

both are parts of the process of river valley formation, the first is a single factor within the second. In addition, most of the alternatives are too long and each one is not only plausible but is partially correct depending on how the pupil decides to interpret the stem.

---------- // ----------

Major sources:

(1) Betrand Arthur and Cebula Joseph P. **Tests, Measurement, and Evaluation: A Development Approach**. London: Addison-Wesley, 1980

(2) Gronlund, Norman E. **Measurement and Evaluation in Teaching**. New York: MacMillan, 1976.

(3) Nelson M.J. **Tests and Measurements in Elementary Education.** New York: The Gordon Company, 1939.

PART TWO - in the **NEXT ISSUE!**

---------- // ----------