CHAPTER **10** Classroom Assessment

LEARNING OUTCOMES

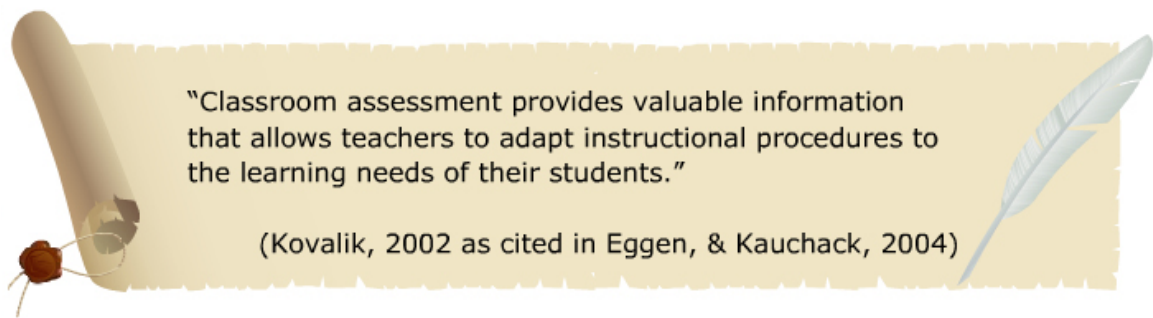After studying this chapter, you should be able to:

1. Define assessment;

2. Explain the basic concepts in assessment;

3. Explain how to plan for assessment;

4. Describe types of assessment in the classroom;

5. Explain what is teacher-made tests;

6. Describe what is standardized tests; and

7. Explain what is authentic assessment.

INTRODUCTION

One of the most basic and difficult task that teachers face in their work is the process of assessment. Classroom assessment includes all the process involved in making decisions about students learning progress. It includes the observation of students' written work, their answers to questions in class, and performance on teacher-made and standardized tests.

According to (Koyalik, 2002 as cited in Eggen & Kauchak, 2004):

"Classroom assessment provides valuable information that allows teachers to adapt instructional procedures to the learning needs of their students."

(Kovalik, 2002 as cited in Eggen, & Kauchack, 2004)

It facilitates teachers in decision making about learning progress through systematic information gathering.
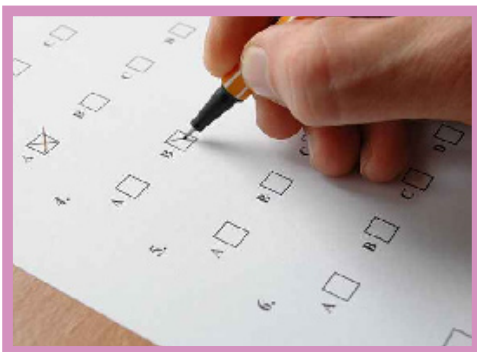
Besides that, assessment also accomplishes two other important goals; increasing learning and increasing motivation.

The relationship between learning and assessment is very strong. Students learn more in classes where assessment is an integral part of instruction than in those where it isn't. Brief assessment that provides frequent feedback about learning progress is more effective than long, infrequent ones, like once-a-term tests.

10.1 DEFINITION OF MEASUREMENT, EVALUATION, ASSESSMENT AND TEST

Assessment, evaluation, measurement and test: all four terms have to do with the systematic process of collecting data and analysing those data to make decisions. These four terms may all mean the same thing, but there are some important differences among them, especially to teachers who engage in these processes in their classrooms. However, some experts give more narrow definitions of these terms.

Below are the four narrow definitions of the terms:

Measurement	The process of gathering information about learning. it is the process of quantifying the degree to which someone or something possesses a given characteristic, quality or feature. The purpose of measurement is to produce quantitative data, such as test scores, numerical rating on assigned projects, height, weight, visual acuity and number of sit-ups completed in one minute. The numerical data generated in the measurement process are typically used in the evaluation of something, for example students' achievement.
Evaluation	The process of making decisions on the basis of measurements. For example, a school district may evaluate alternative approaches to teaching reading to determine which one work best with the students in serves. Teachers evaluate how well or to what extent their students are achieving instructional outcomes. Teachers also evaluate students at the end of the year to determine whether they are suitably prepared for the work expected at the next grade level. Good evaluation requires data gathering for informed decision making.
Assessment	The purposeful process in the classroom to collect data, both qualitative and quantitative. Hence, it is a broader term than measurement. For teachers, the purpose of assessment is usually to make decisions about students either as a group or individually. For example, teachers diagnose difficulties, verify learning after instruction, identify prerequisite learning and determine where to start in a learning sequence based on students already know. In the classroom, assessment considers students performances on tasks in a variety of settings and contexts. Assessment also may be intrusive or not, in other words, students may know that they are being assessed or it may be seamless with instruction, perhaps

	perceived bay the students as nothing more than a chance to practice what they have learned. Data collection in assessment is accomplished in a variety of ways such as observation of behavior, presentation of tasks for students to perform mentally and to record their response in writing or examination of products produced by the students.
Test	A question or a task or a series of such, designed to elicit some predetermined behavior from the person being tested. The terms assessment and test are usually used interchangeably by many people. However, the word test implies a paper-and – pencil instrument, administered under pre-specified conditions that are consistent across students. Its traditional usage by teachers has implied a written series of tasks to which students responded in writing (for example essay or short-answer items) or marked their response choices with a pen or pencil (for example true-false, multiple choice, or matching items) (Gallagher, 1998).

SELF-CHECK

1. What is the meaning of assessment?

2. Can you differentiate the meaning of assessment with the meaning of measurement, evaluation and test?

10.2   BASIC CONCEPTS IN ASSESSMENT

This section describes how teachers can apply the principles of behaviourism in the classroom. It is divided into three subsections according to the perspectives discussed earlier that are classical conditioning, operant conditioning and social learning theory.

10.2.1   Reliability

Reliability   refers to the extent to which assessments are consistent. Just as we enjoy having reliable cars (cars that start every time we need them), we strive to have reliable, consistent instruments to measure students' achievement. Another way to think of reliability is to imagine a kitchen scale. Let us look at the examples and the similarities in Table 10.1 below.

Table 10.1: Example and similarity of 'Reliabity'

Example	Similarity
If you weigh five pounds of potatoes in the morning, and the scale is reliable, the same scale should register five pounds for the potatoes an hour later (unless, of course, you peel and cook them).	Instruments such as classroom tests and national standardised exams should be reliable – it should not make any difference whether a student takes the assessment in the morning or afternoon; one day or the next.
If you create a quiz to measure students' ability to solve quadratic equations, you should be able to assume that if a student gets an item correct, he or she will also get other, similar items correct.	Another measure of reliability is the internal consistency of the items.

10.2.2 Validity

Validity refers to the accuracy of an assessment; whether or not it measures what it is supposed to measure. Even if a test is reliable, it may not provide a valid measure. Consider a situation below.

Situation:

Let us imagine a bathroom scale that consistently tells you that you weigh 130 pounds. The reliability (consistency) of this scale is very good, but it is not accurate (valid) because you actually weigh 145 pounds.

According to Hogan, 2007,

"Since teachers, parents and school districts make decisions about students based on assessments (such as grades, promotions, and graduation), the validity inferred from the assessmentis essential, even more crucial than the reliability. Also, if a test is valid, it is almost always reliable" .

(Hogan, 2007)

So,

Does all this talk about validity and reliability mean you need to conduct statistical analyses on your classroom quizzes?

No, it does not.
(Although you may, on occasion, want to ask one of your peers to verify the content validity of your major assessments.)
However, you should be aware of the basic tenets of validity and reliability as you construct your classroom assessments, and you should be able to help parents interpret scores for the standardised exams.

1. Your school district is looking for an assessment instrument to measure reading ability. They have narrowed the selection to two possibilities. Test A provides data indicating that it has high validity, but there is no information about its reliability. Test B provides data indicating that it has high reliability, but there is no information about its validity. Which test would you recommend? Why?

10.3   PLANNING FOR THE ASSESSMENT

The classroom assessment process assumes that students need to receive feedback early and often, that they need to evaluate the quality of their own learning, and that they can help the teacher improve the strength of instruction.

The basic steps in the classroom assessment process are setting targets and writing objectives, choose an assessment items and technique, administering assessments and analyse the data and share the results with students.

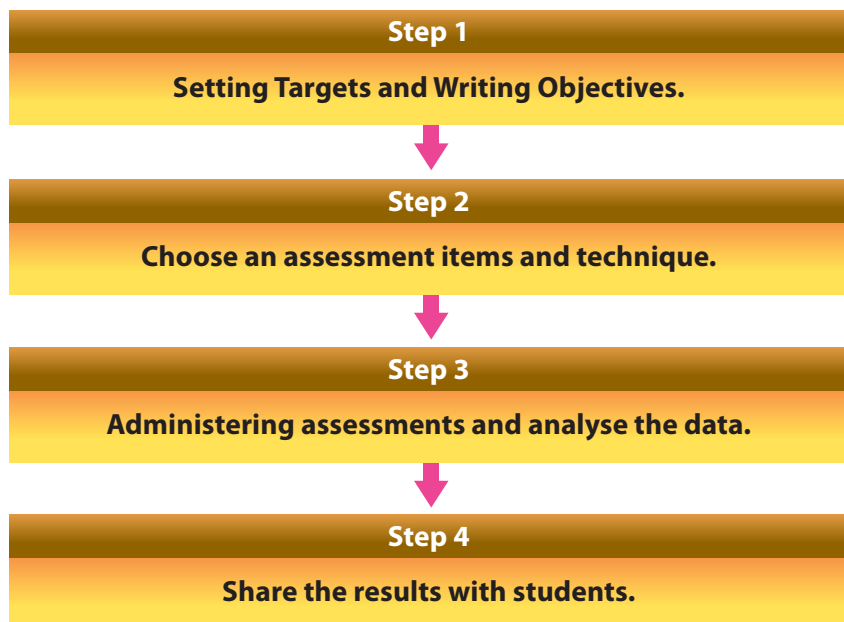Figure 10.1 shows 4 steps of classroom assessment process.

Step 1
Setting Targets and Writing Objectives.

Step 2
Choose an assessment items and technique.

Step 3
Administering assessments and analyse the data.

Step 4
Share the results with students.

Figure 10.1: The four steps of a classroom's assessment process.

10.3.1   Setting Targets and Writing Objectives

Setting clear and achievable targets is the starting point for creating assessments. In other words, you need to determine what exactly your students should know or be able to do. If you do not set clear targets, you will never know if the instruction and experience in the classroom resulted in a "bull's-eye" or if they miss the mark completely.

There are many areas and types of achievement that are targeted in schools, including knowledge, reasoning, performance, product development and attitudes. As indicated in Table 10.2, there are many ways to assess each of these areas.

Table 10.2: Target area and the possible ways to assess it

Target Area	Example Target Behaviour	Possible Assessments
Knowledge	Spell words correctly	Quizzes, essays, questioning
Reasoning	Solve math problems	Essays, observations
Performance	Speak foreign language	Observations, rubrics
Product Development	Create a web page	Rubrics
Attitudes	Positive attitudes	Surveys, observations

Using these general target areas, teachers create specific classroom objectives that are based on state/district standards and benchmarks. Well-written objectives are made up of three elements which is conditions, behaviour and criterion as shown in Figure 10.2.
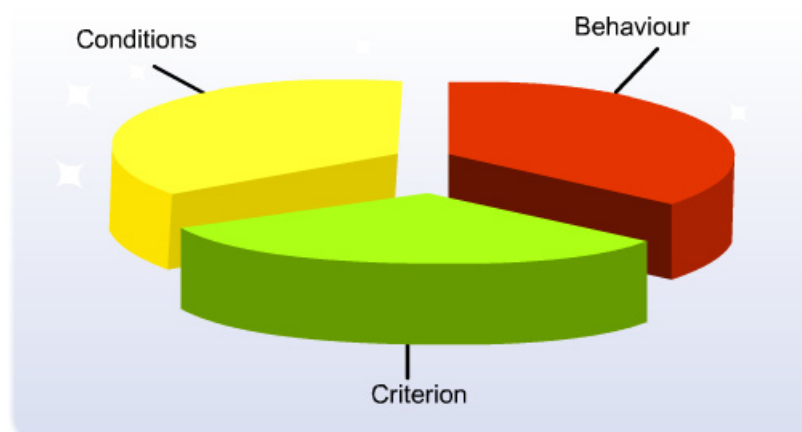

Figure 10.2: The three elements that made up well-written objectives.

The descriptions for these 3 elements that made up well-written are explained in Table 10.3.

Table 10.3: The definition and examples of the three elements.

Element	Definition	Example
Conditions	The materials that will be available (or unavailable) when the objective is assessed. It generally states what the student will be *given* or *not given.*	● Without the use of a calculator... ● Given a map of Europe... ● Given 12 double-digit numbers...
Behaviour	A verb that describes an observable activity such as what the student will do. The behaviour is generally stated as an action verb.	● Solve ● Compare ● List ● Explain ● Evaluate ● Identify
Criterion (also referred to as Degree)	The standard that is used to measure whether or not the objective has been achieved. The criteria might be stated as a percentage (80% correct), a time limit (within five minutes), or another measure of mastery	An objective might be: "Given a list of 20 states (condition),the student will identify (behaviour) at least 15 of the corresponding state capitals (criteria). " After the objectives are written, it is relatively easy to create a corresponding assessment item. It isimportant to ensure that assessments are consistent with the objectives and instruction. This seems obvious, but because tests are usually prepared some time after instruction is completed, it often does not occur. For example, a topic may be discussed in class at applied level,but test items occur at a knowledge level.

10.3.2   Choose an Assessment Items and Technique

Once teachers have determined the purpose of a test and identified those behaviours that are critical for mastery, they should give thought to those items and techniques that will best serve their purpose.

What type of test items should be used?

There are many types of items of assessment that teachers can choose as performance items (such as project paper), interviews, objectives items and essay items.

However, most teachers will be concerned with either objective or essay items. Table 10.4 shows some distinguishing characteristics between objective and essay items.

Table 10.4: The distinguishing characteristics of objectives and essay test item.

Objective Test Items		Essay Test Items	
Supply items	**Selection items**	**Extended response**	**Restricted response**
Require students to give an answer.	Require students to choose from among several alternatives. More highly structured and restrict the types of response students can make.	Furnishes students with complete freedom to make any kind of response they choose.	Asks for specific information, thus restricting students' responses (Elliott, Kratochwill, Cook & Travers, 2000).

10.3.3 Administering Assessments

When teachers administer tests and quizzes, they want to create conditions that optimize students' performance to ensure that assessment results accurately reflect what students know and can do. Table 10.5 shows what should be done in administering the test.

Table 10.5: Three step that should be done in administering the test.

1	Arrange the environment to be comfortable, free from distractions and similar to the way it was when students learn the content. Distraction can depress test performance, particularly in young and low-ability students.
2	Give the students precise directions about how to take the test, how the teacher will collect the papers, and how students should spend their time afterward. These directions help maintain order and prevent distraction for late finishing students.
3	Carefully monitor the test during the entire time students work on it. This will encourage students who become lost or distracted and also discourage cheating. According to Eggen and Kauchack, 2004: "Teacher monitoring during test also helps students learn to monitor their own test-taking behaviour." (Eggen, & Kauchack, 2004)

10.3.4 Analyse the Data and Share the Results with Students

Assessment effort does not end with administering the test. The teachers have to score it and return it to the students, discuss the results and provide students with feedback as quickly as possible. This process of providing assessment feedback is important for both achievement and motivation. Feedback allows students to correct common misconceptions, and knowledge of results promotes students' motivation.

Virtually, all teachers provide feedback after a test, and many take up to half a class period to do so. Students' motivation is high and many teachers believe that students learn more in these sessions than they do in the original instruction. According to Bridgeman, as citied in Eggen & kauchak,2004,

"Research showed that students who were told that they did well on a test perform better on a subsequent measure than those who were told that they did poorly, even though the two groups did equally well on the first test."

(Bridgeman, 1974 as cited in Eggen & Kauchack, 2004)

SELF-CHECK

1. What are the four steps in planning a classroom assessment?

2. As a teacher you may plan to make an assessment in your classroom. What are the things that you should consider?

3. Well written objectives are made up of three elements. Explain each of those elements.

4. Suppose that you were administrating test in your classroom. How would you do to ensure that the test goes smoothly and that the assessment results accurately reflect what students know and can do?

10.4   TYPES OF ASSESSMENT IN THE CLASSROOM

Almost everyone knows about the types of tests typically encountered in school. There are final exams, midterm exams, end-of unit tests, pop quizzes and so on. All of those tests have one thing in common. They represent teacher's attempt to get a fix on how much the students have learned. More accurately, such tests are employed to determine a student's status with respect to the knowledge and skills that the teacher is attempting to promote. If teachers are reasonably sure about what their students currently know, then the teacher can more accurately tailor instructional activities to what students need to know.

The types of assessments such as the quizzes and examinations that most of us took in school, have historically been paper-and pencil instruments. However, in recent years, educators have been urged to broaden their conception of testing so that students' status is determined via a wider variety of measuring devices.

10.4.1   Formative and Summative

Formative and summative assessments are the two types of assessments that are always used by teachers. Now let us look what are formative and summative assessments and then the common

things between them in Table 10.5.

a) **Formative assessments** are on-going assessments, reviews, and observations in a classroom. Teachers use formative assessment to improve instructional methods and student's feedback throughout the teaching and learning process. For example, if a teacher observes that some students do not grasp a concept, she or he can design a review activity or use a different instructional strategy. Likewise, students can monitor their progress with periodic quizzes and performance tasks.
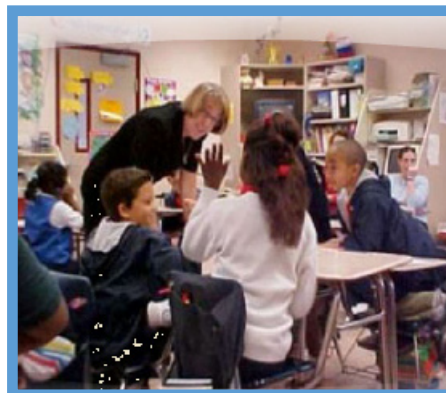
The results of formative assessments are used to modify and validate instruction. Formative assessment is generally carried out throughout a course or project. Formative assessment also referred to as 'educative assessment', is used to aid learning. In educational setting, formative assessment could be a teacher (or peer) or the learner, providing feedback on a student's work and would not necessarily be used for grading purposes.

b) **Summative assessments** are typically used to evaluate the effectiveness of instructional programs and services at the end of an academic year or at a pre-determined time. The goal of summative assessments is to make a judgment of student competency, after an instructional phase is complete. For example, in Malaysia, the final examination is administered once a year. It is a summative assessment to determine each student's ability at pre-determined points in time.

Summative evaluations are used to determine if students have mastered specific competencies and to identify instructional areas that need additional attention. Summative assessment is generally carried out at the end of a course or project. In an educational setting, summative assessments are typically used to assign students a course grade.

Table 10.5 shows formative and summative assessments that are common in schools.

Table 10.6 : Common formative and summative assessments in schools.

Formative Assessments	Summative Assessments
Anecdotal records	Final exams
Quizzes and essays	National tests (UPSR, PMR, SPM)
Diagnostic tests	Entrance exams

Summative and formative assessments are often referred to, in a learning context, as assessment of learning and assessment for learning respectively. Assessment of learning is generally summative in nature and intended to measure learning outcomes and report those outcomes to students, parents and administrators. Assessment of learning generally occurs at the conclusion of a class, course, semester or academic year. Assessment for learning is generally formative in nature and is used by teachers to consider approaches to teaching and the next steps for individual learners and the class.

10.4.2 Objective and Subjective

Assessment (either summative or formative) is often categorized as either objective or subjective.
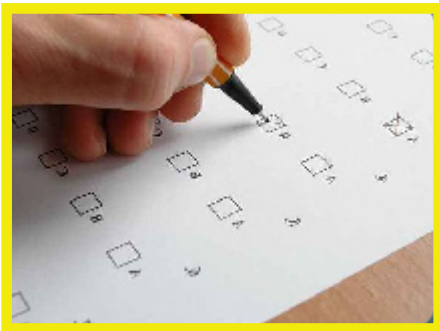
Objective assessment is a form of questioning which has a single correct answer. It is also known as Selected-Response Items.

A scoring key for correct responses is created and can be applied by an examiner or by a computer. The scoring is easy, objective and reliable. The task is highly structured and clear where it can measure both simple and complex learning outcomes. However, constructing good items is time consuming and it is ineffective to measure some types of problem-solving items.
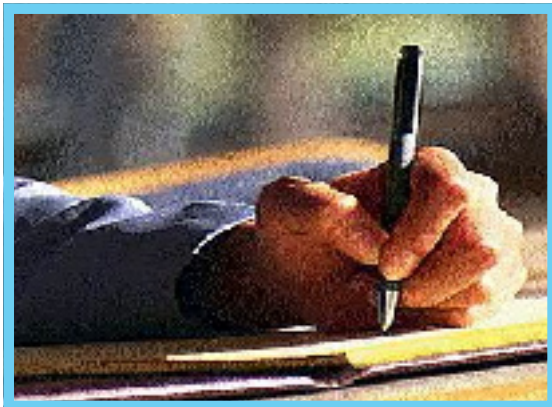
Subjective assessment is a form of questioning which may have more than one correct answer (or more than one way of expressing the correct answer). It is known as Constructed-Response Items.

It requires students to write out information rather than select a response from a menu. In scoring, many constructed-response items require judgment on the part of examiner. It can measure the highest level of learning outcomes such as analysis, synthesis and evaluation

and the integration and application of ideas can be emphasized. In term of preparation, the essay types questions can be prepared in less time compared to selection-types format.

However, the disadvantage is that scoring is time consuming, subjective and possibly unreliable. There are various types of objective and subjective questions. Objective question types include true/false answers, multiple-choice, multiple-responses and matching questions. Subjective questions include extended-response questions, restricted-response questions and essays. Objective assessment is well suited to the increasingly popular computerized or online assessment format.

Table 10.7 shows various types of objective and subjective assessments.

Table 10.7 : Various types of objective and subjective assessments.

Objective Assessments	Subjective Assessments
True/False items	Extended-response items
Multiple choice items	Restricted-response items
Multiple-responses items	Essays
Matching items	

Some have argued that the distinction between objective and subjective assessments is neither useful nor accurate because, in reality, there is no such thing as 'objective' assessment. In fact, all assessments are created with inherent biases built into decisions about relevant subject matter and content, as well as cultural (class, ethnic, and gender) biases.

10.4.3	Bases of Comparison

Test results can be compared against an established criterion, or against the performance of other students, or against the previous performance.

a) **Criterion-referenced assessment**

Typically, using a *criterion-referenced test*, as the name implies, occurs when candidates are measured against defined (and objective) criteria. Criterion-referenced assessment is often, but not always, used to establish a person's competence (whether she or he can do something). The best known example of criterion-referenced assessment is the driving test, when learner drivers are measured against a range of explicit criteria (such as 'Not endangering other road users').

b) Norm-referenced assessment

Typically, using a *norm-referenced test*, is not measured against defined criteria. This type of assessment is relative to the student undertaking the assessment. It is effectively a way of comparing students. The IQ test is the best known example of norm-referenced assessment. Many entrance tests (to prestigious schools or universities) are norm-referenced, permitting a fixed proportion of students to pass ('passing' in this context means being accepted into the school or university rather than an explicit level of ability). This means that standards may vary from year to year, depending on the quality of the cohort, criterion-referenced assessment on the other hand does not vary from year to year (unless the criteria change).

10.4.4 Informal and Formal

Assessment can be either formal or informal. Table 10.7 shows the diference between formal and informal assessment.

Table 10.8 : Formal and informal assessment.

Formal Assessment	Informal Assessment
● Usually implicates a written document, such as test, quiz or paper.	● Usually occurs in a more casual manner and may include observation, inventories, checklists, rating scales, rubrics, performanceand portfolio assessments, participant, peer and self evaluation and discussion.
● Given a numerical score or grade based on student's performance.	● Does not contribute to a student's final grade

10.4.5 Internal and External Assessment

Internal assessment is set and marked by the school (i.e. teachers). Students get the mark and feedback regarding the assessment. External assessment is set by the governing body, and is marked by non-biased personnel. With external assessment, students only receive a mark. Therefore, they have no idea how they actually performed (i.e. what bits they answered correctly.)

SELF-CHECK

1. Identify at least three differences between formative and summative assessment?

2. What are the strengths of multiple-choice items compared to essay items?

3. Informal assessments are often unreliable, yet they are still important in classrooms. Explain why this is the case, and defend your explanation with example.

10.5   TEACHER-MADE TESTS

Individuals interested in constructing a test are confronted with challenges concerning what to assess, how to assess it, and whether they measure it in a reliable and valid manner. These are fundamental challenges to teachers who construct their own tests and to professionals who design standardised tests. Consequently, these topics deserve our attention.

10.5.1   Teacher-made Paper-and-pencil Tests

Although teachers use many techniques in evaluating students, probably the most popular is the written paper-and-pencil test that they themselves construct. These usually consist of essay or multiple-choice items. The multiple choice, paper-and-pencil test is probably the most frequently used test, with other types such as true/false, essay and performance tests.

"Good multiple choice items are difficult to prepare but can be scored easily and objectively. Essay tests, on the other hand, are relatively easy to prepare but extremely difficult to score."

(Elliott, Kratochwill, Cook, & Travers, 2000)

10.5.2   Teacher-made Performance and Product Assessment

To assess some instructional outcomes or evaluate certain areas of students' learning, using paper-and-pencil tests is inappropriate. For example, teacher cannot determine how well a student can type a letter with a multiple choice test. Sometimes, a paper-and-pencil assessment will do, but it is not always the best method, there are instructional outcomes for which other assessment option works better.

For certain assessment situation, observation is clearly the most appropriate approach. For example, attitudes can be assessed by asking persons to respond in writing to a number of questions. But a person may say he is a good sport and then break the rules to win or walk away pouting. In such cases, more accurate information could be obtained by actually watching him compete.

Observation is used as a method of assessment when actions speak louder than words. In observation assessment, teachers gather data not by asking for information but by watching closely. The students being observed, usually does not write anything as she would on a paper-and-pencil test. Instead, the student performs some actions and her behaviour is observed and recorded by the teacher.
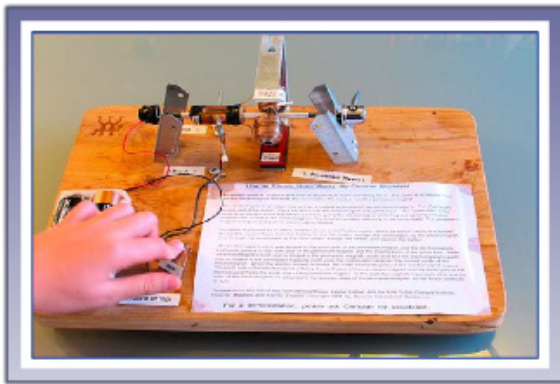
Performance assessments include what is commonly thought of as students' actual performance such as oral presentation, dance, music, art and physical education. In addition to assessing student performance, student products may also be assessed. Book report, papers, dioramas, science fair projects, art work and portfolios are all examples of tangible output that can be assessed on a number of dimensions depending on the teacher's objectives for the class. Even a student-composed answer to an essay question on a written test can be considered a product.

According to Gallagher:

> "The teacher decides the important dimensions and criteria for success the product accordingly."
>
> **(Gallagher, 1998)**

Table 10.9 shows the two performance assessments which is observation and student products.

Table 10.9 : Examples of the two performance assessments.

Observation	Student Products
Oral presentation	Book report / papers
Dance	Dioramas
Music	Science fair projects / experiment
Art	Art works
Physical education	Portfolios

10.5.3 Teacher-made Rating Scales and Checklist

Rating scales and checklists are instruments teachers use to help with data-gathering when assessing either a performance or a product. A **rating scale** is typically an instrument with a number of items related to a given variable, each item representing a continuum of categories between two extremes, usually with a number of points along the continuum highlighted in some way. Persons responding to the items place a mark to indicate their position to each item.

A **checklist** enumerates a number of behaviours or features that constitute a procedure or product. When a procedure is involved, the steps are typically listed in the desired order. According to Gallagher,

> "The person completing the checklist indicates whether a given behaviour or feature occurred or is present (Gallagher, 1998)."
>
> **(Gallagher, 1998)**

Teachers are not psychometricians. Thus, the assessment instruments (such as tests and observation checklists teachers produce) are not likely to be formally validated, field-tested and revised. Larger budgets than those teachers normally have at hand are required to produce assessments that meet such rigorous requirements. For these assessments, teachers usually turn to commercial test publishers.

SELF-CHECK

1. Suppose that you have been told that as part of your teacher certification process, there will be a performance assessment. What do you think is the best performance assessment of your teaching skills?

2. What are some features of rating scale? How it is differ from checklist?

3. If you are a teacher, choose one suitable assessment that can be used to measure your students' performance? Explain why you do think that type of assessment is suitable to measure your target area?

10.6 STANDARDISED TESTS

Standardised tests are no better as assessment tools than teacher-made tests. Standardised tests are better suited to large scale data collection and when uniform comparisons across students are crucial.

According to (Airasian, 1997 as cited in Tan, Parsons, Hinson, & Sardo-Brown, 2003):

"Standardised tests are intended to be administered, scored, and interpreted in the same way for all test takers, regardless of where or when they are assess."

(Airasian, 1997 as cited in Tan, Parsons, Hinson, & Sardo-Brown, 2003)

These tests can be administered individually or to groups and provide students with feedback. It often served the bureaucratic needs of educational leaders. It is also used to inform admission and selection procedures, to sort and identify the special needs of students, and to provide accountability information about the efficacy of schooling at all levels. Many standardised tests, therefore, are norm-referenced by design, as they are created to make comparison between students along specified measurement. Most standardised tests measure aptitude and achievement.

10.6.1 Standardised Achievement Test

There are some characteristics of standardised achievement test as discussed follows.

The tests aim to measure attainment of objectives in school-based curricula. It explicitly try to gauge skills and knowledge developed as a result of specific instruction. This is the principal that differentiate it with the 'ability' test.

Most of the standardised tests use multiple-choice items at least to a substantial degree. Some of the tests are entirely multiple choice. Many use mixture of multiple-choice and constructed-response items. Use of constructed-response items is increasing modestly. However, multiple-choice items still predominate in most areas.

Developers of standardized achievement tests typically pay considerable attention to the technical characteristics desired for tests. They usually have reliability data, item analysis and other such technical information for these tests. Such information usually appears in a technical manual or other formal report for the test. In addition, the pre-publication research typically includes professional reviews for cultural, racial, ethnic, and gender bias.

The interpretation of scores on standardised achievement tests relies on use of large-usually been quite good. When the norm is not national in scope, it is at least based on a reasonably large group considered relevant for the purposes of the test. For example, the norm may be based on all grade 4 students in a state or all candidates for a certain type of license.

10.6.2 Standardised Ability Test

Human beings possess and display a wide range of abilities. Some are easy to identify, for example athletic ability, musical talent and artistic flair. Some abilities are more difficult to recognise.Traditionally, the type of ability of most interest in the educational circles has been the mental ability. Most people believe that mental ability has something to do with success in school. However, it is not the only important factor to succeed in school.

Effort, motivation, concentration and other variables are also important. But, it seems that human beings have a level of mental ability that makes school learning more or less difficult. The mental ability tests used in schools fall into two general classifications. They are individual administered tests and group administered tests.

(a) Individual administered tests

A teacher would not ordinarily administer such tests. However, teachers will encounter reports of scores from these tests. Thus, teachers need to be familiar with them. It is expensive and time-consuming to give these tests. Usually, they are administered only to students with special needs. These tests play a key role in the identification of learning disabilities, mental retardation and other such conditions. They may also be used in the selection of students for gifted programs. Most students will complete school without ever taking one of these tests.

The most obvious feature of the group test is that, as suggested by the name of this category, it can be administered to a group all at one time. The most typical arrangement would be administration to a classroom of students. However, with appropriate spacing and proctoring, these tests can be administered to hundreds of people at once.

Administration of these tests does not require a specialised training needed to administer the individual tests. Many teachers will administer one of these group tests to their students. Items appearing in group tests are very similar in many ways to the items appearing

in individually administered tests.

According to Hogan, 2007:

> "The group tests essentially tried to duplicate, as far as possible, the individual tests but in format suitable for group administration."
>
> **(Hogan, 2007)**

SELF-CHECK

1. What are some of the features of standardised tests?

2. Identify at least 3 examples of ability that can be measured and what type of ability gain most interest in educational circles?

3. Most standardised tests measure achievement. List down some of the criteria of achievement test?

10.6.3 Standardised Aptitude Test

Aptitude tests are used to predict what students can learn. Aptitude tests do not measure innate capacity or learning potential directly, rather they measure performance based on learning abilities. Intelligence tests are perhaps the best example of aptitude tests commonly used in schools. It is interesting to note some differences between aptitude and achievement tests. According to Elliot, Kratochwill, Cook and Traverse,

> "An aptitude test predict and individual's performance in a certain task or in particular job by sampling the cumulative effect on the individual of many experiences in daily living, including specific educational experience. Aptitude tests measure only innate capacity, while achievement tests measure only the effects of learning."
>
> **(Elliott, Kratochwill, Cook, & Travers, 2000)**

10.6.4 Standardised Personality Test

There are several types of personality measures. Let us review the three types of personality measures; namely self report inventories, projective techniques and behaviour rating scales as shown in in Figure 10.3.
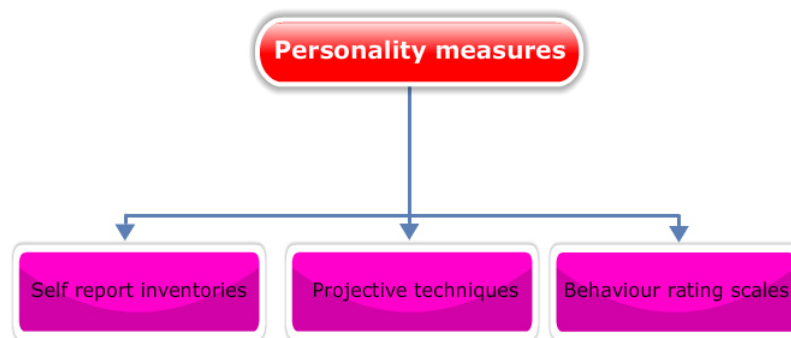
Figure 10.3: The three types of personality measures.

Many of these tests concentrate on clinical applications that fall outside the realm of ordinary educational assessment. Therefore, we provide just an overview of these categories, with special reference to the types of measures a teacher might encounter.

(a) **Self report inventories**

The most widely used type of personality test. This type of test consists of simple statement to which a person responds True or False, Yes or No or similarly simple options. The statements are very simple and the response is also simple. They are sometimes called objective inventories, because they can objectively scored.

(b) **Projective techniques**

Present a person with an ambiguous or innocuous stimulus and a simple instruction about how to respond. The person constructs a response with near maximum freedom. The response reveals something about the individual's personality, motivations and inner dynamics. The use of these techniques requires advance training and supervised experience. Even then their value is controversial. They are widely used by professionals such as school of psychologist. Therefore, although teachers would never administer these instruments, they do need to have some familiarity with them. The classic, well-known projective technique is the Rorschach Inkblot Test and Thematic Apperception Test (TAT).

(c) Behaviour rating scales

Now routinely used in schools for determination of such conditions; attention disorder, hyperactivity, depression, assorted emotional problems. It has two essential features, first, someone other than the person being evaluated completes the rating such as teacher, parents or other caregivers. Second, as suggested by the title, lists specific behaviours. The person completing the form indicates frequency of observing the behaviour.

Hogan, 2007 stated that:

"The descriptor usually short: 1 to 3 words. The ratings are made on 3 to 5 point scale, typically ranging from "never" to "always" or "definitely Not true" to "Definitely true".

(Elliott, Kratochwill, Cook, & Travers, 2000)

10.7	Authentic Assessment

Authentic assessment is a kind of assessment that directly measures students' performance through real-life tasks or product. This alternative assessment includes the following tasks or products such as creating an original piece of artwork, writing a paper, delivering speech and so on. Often teachers who use authentic assessment are interested not only in the products of learning but also in the processes that students use to prepare such products. Thus, portfolio of writing samples may be used to chart the development of students' writing skills over time as they relate to the production of a final editorial. In some cases, teachers may videotape students practising their delivery of speech on successive occasions to document their growth in the development of final version.

The authentic assessment has several advantages compared to the traditional assessment. These advantage of authentic assessment are as explained follows.

They use real-world applications in which students are asked to be active participants in performing, creating or producing something.

They are more likely than traditional assessments to call upon higher-order thinking or problem solving skills.

They ensure that the students are actively involved in constructing understanding.

There are two most popular forms of authentic assessment. The two forms of authentic assessment are performance assessment and portfolio assessment.

10.7.1   Performance Assessment

Performance assessment of higher-level thinking often emphasis 'doing', open-ended activities for which there is no one correct answer. The tasks are sometimes realistic and many, but not all, performance assessments are authentic. Evaluating performance often includes direct methods of evaluation, self-assessment, assessment of group performance as well as individual performance and an extended period of time.

There are four main guidelines in using performance assessment as listed in Figure 10.4.
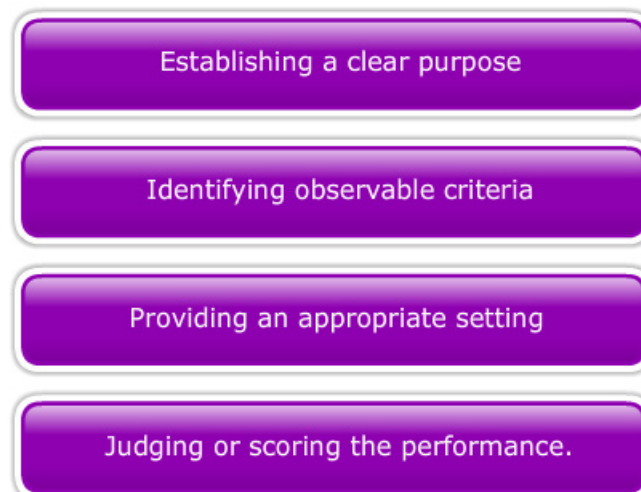
Establishing a clear purpose

Identifying observable criteria

Providing an appropriate setting

Judging or scoring the performance.

Figure 10.4: Four main guidelines in using performance assessments.

10.7.2   Portfolio Assessment

A portfolio is a systematic and organised collection of a students' work compiled by students and teachers that are reviewed against preset criteria to judge a student or program. Four classes of evidence can be included; artefacts, reproductions, attestations, and production. The collections consist of the products of learning such as a videotape, a piece of artwork, a journal entry or an essay and the portfolio items are different according to the content area.

A portfolio provides a tangible evidence of accomplishment and skills that must be updated as the person grow and change.

- ✔ Scoring and judging the portfolio.
- ✔ Setting criteria for evaluation.
- ✔ Reviewing the portfolio with students.
- ✔ Involving students in decisions about it.
- ✔ Establishing the portfolio's purpose.

Two broad types of purposes of portfolio are to document growth through a growth portfolio and to showcase the students' most outstanding work through a best-work portfolio. The strengths of learning portfolio such as capturing the complexity and completeness of the students' work and accomplishments, as well as encouraging students' decision-making and self-reflection. The weaknesses of learning portfolio, such as the time required to coordinate and evaluate them and the difficulty in evaluating them.

10.7.3 Six Principles of Authentic Assessment

They are 6 principles of Authentic Assessment. The description for those principles are as follows.

1	Authentic assessment is continuous, informing every aspect of instruction and curriculum building. As they engage in authentic assessment, teachers discover and learn what to teach as well as how and when to teach them.
2	Authentic assessment is an integral part of the curriculum. Children are assessed while they are involved with classroom learning experiences, not just before or after a unit through pre or post tests.
3	Authentic assessment is developmentally and culturally appropriate.
4	Authentic assessment focuses on students' strengths. Teachers assess what students can do, what they know, and how they can use what they know to learn.
5	Authentic assessment recognizes that the most important evaluation is self evaluation. Students and teachers need to understand why they are doing what they are doing so that they may have some sense of their own success and growth.
6	Authentic assessment invites active collaboration between teachers, students and parents work together to reflect and assess learning (Bridges, 1995).

SELF-CHECK

1. What makes an assessment "authentic"?

2. How can portfolio be used in assessment?

3. What is the differences between performance and portfolio assessment?

SUMMARY

- The terms assessment, evaluation, measurement and test may all mean the same thing, but there are some important differences among them.

- The starting point for creating assessments includes setting clear and achievable targets and writing objectives.

- Reliability and validity are two concepts that are important in assessment. *Reliability* refers to the extent to which assessments are consistent. *Validity* refers to the accuracy of an assessment; whether or not it measures what it is supposed to measure.

- There are four basic steps in planning for assessment;
  1) choose a learning goal to assess
  2) choose an assessment items and technique
  3) apply the technique and
  4) analyze the data and share the results with students

- There are several types of assessment in the classroom. In recent years, the conception of testing has broadened up so that students' status is determined via variety of measuring devices. Those types of assessments can be divided into formative and summative, criterion-referenced assessment & norm-referenced assessment, formal & informal, internal & external.

- Teachers who construct their own tests are confronted with challenges concerning what to assess, how to assess it, and whether they measure it in a reliable and valid manner. There are several types of test that teachers usually construct such as teacher-made paper-and-pencil tests, teacher-made performance and product assessment and teacher-made rating scales and checklists.

- Standardized tests is a type of assessment that suited to large scale data collection and when they are intended to be administered, scored, and interpreted in the same way for all test takers, regardless of where or when they are assessed. There are variety of standardized tests such as achievement tests, ability tests, personality tests, and aptitude tests.

KEY TERMS AND CONCEPTS

Terms	Definitions
Ability Test	**The tests aim to measure a wide range of human beings' abilities such as athletic ability, musical talent, artistic flair and mental ability, the type of ability of most interest in educational circles.**

Achievement Test	**The tests aim to measure attainment of objectives in school-based curricula. It explicitly try to gauge skills and knowledge developed as a result of specific instruction.**
Aptitude tests	**The tests are used to predict what students can learn. Aptitude tests do not measure innate capacity or learning potential directly, rather they measure performance based on learning abilities. For example = Intelligence tests.**
Assessment	**The purposeful process in the classroom to collect data, both qualitativeand quantitative; in the classroom, assessment considers students performances ontasks in a variety of settings and contexts.**
Assessment feedback	**The process where teachers score the test and return it to the students, discuss the results and provide students with feedback as quickly as possible to allows students to correct common misconceptions, and knowledge of results promotes students motivation.**
Assessment for learning	**Assessment that is generally formative in nature and is used by teachers to consider approaches to teaching and next steps for individual learners and the class.**
Assessment for learning	**Assessment that is generally summative in nature and used by teachers to measure learning outcomes and report those outcomes to students, parents, and administrators. It is generally occurs at the conclusion of a class, course, semester, or academic year.**
Authentic assessment	**A kind of assessments that directly measure student performance through real-life tasks or product such as creating an original piece of artwork, writing a paper, delivering speech and so on.**

Checklist	**Enumerates a number of behaviors or features that constitute a procedure or product. When a procedure is involved, the steps are typically listed in the desired order.**
Constructed-Response Items	**It is known as subjective assessment. It is a form of questioning which may have more than one correct answer (or more than one way of expressing the correct answer). It requires students to write out information rather than select a response from a menu.**
Criterion-referenced assessment	**Assessment that is used when candidates are measured against defined (and objective) criteria. It is often, but not always, used to establish a person's competence (whether she or he can do something) – example = driving test.**
Evaluation	**The process of making decisions on the basis of measurements.**
Formal assessment	**Assessment that is given a numerical score or grade based on student performance, usually implicates a written document, such as a test, quiz, or paper.**
Formative assessments	**The on-going assessments, reviews, and observations in a classroom; assessment during the course of instruction rather than after it is completed.**
Informal assessment	**Assessment that is usually occurs in a more casual manner and may include observation, inventories, checklists, rating scales, rubrics, performance and portfolio assessments, participation, peer and self evaluation, and discussion. It does not contribute to a student's final grade.**
Measurement	**The process of gathering information about learning. It is the process of quantifying the degree to which someone or something possesses a givencharacteristic, quality or feature.**

Norm-referenced assessment	**Assessment that is not measured against defined criteria. It is relative to the student body undertaking the assessment. It is effectively a way of comparing students–example = IQ test.**
Observation	**A method of assessment where teachers gather data not by asking for information but by watching closely. The student performs some action and her behavior is observed and recorded by the teacher.**
Performance assessment	**A kind of assessments that often emphasis "doing", open-ended activities for which there is no one correct answer and that may assess higher-level thinking.**
Personality Test	**The tests are used to measure various kind of personality. There are several types of personality measures such as self report inventories, projective techniques and behavior rating scales.**
Portfolio	**A systematic and organized collection of a students' work compiled by students and teachers that are reviewed against preset criteria to judge a student or program.**
Rating scale	**An instrument with a number of items related to a given variable, each item representing a continuum of categories between two extremes, usually with a number of points along the continuum highlighted in some way.**
Reliability	**Refers to the extent to which assessments are consistent.**
Selected-Response Items	**It is also known as objective assessment. It is a form of questioning which has a single correct answer.**
Standardized tests	**The kind of test that are better suited to large scale data collection and when uniform comparisons across students are crucial. It is intended to be administered, scored, and interpreted in the same way for all test takers, regardless of where or when they are assess.**

Summative assessments	Assessment that is used to evaluate the effectiveness of instructional programs and services at the end of an academic year or at a pre-determined time; assessment after instruction is finished.
Test	A question or a task or a series of such, designed to elicit some predetermined behavior from the person being tested.
Validity	Refers to the accuracy of an assessment, whether or not it measures what it is supposed to measure.

ENDNOTES

1. Bridges, L. 1995. Assessment: Continuous learning. California:Stenhouse Publishers.

2. Eggen, P. & Kauchack, D. 2004. Educational psychology:windows on classrooms. 6th ed. New Jersey: Pearson.

3. Elliott, S.N., Kratochwill, T.R., Cook, J.L. & Travers, J.F. 2000. Educational psychology: effective learning. 3rd ed. Boston: McGraw Hill.

4. Gallagher, J.D. 1998. Classroom assessment for teachers. New Jersey:Prentice-Hall Inc.

5. Hogan, T.P. 2007. Educational Assessment: A practical introduction. Danvers:Wiley.

6. Santrock, J.W. 2006. Educational psychology: Classroom update: Preparing for PRAX-ISTM and practice. 2nd ed. New York: Mc Graw Hill.

7. Tan O.S., Parsons, R.D., Hinson, S.L., & Sardo-Brown, D. 2003. Educational psychology: A practitioner-researcher approach. Australia:Thomson.

REVIEW AND DISCUSSION QUESTIONS

1. The processes and tools teachers use to make decisions about their students' progress is best described as

    A. Classroom assessment.
    B. Formal measurement.
    C. Informal measurement.
    D. Formal evaluation.

2. The extent to which teachers gather information about their students that is consistent best describes:

    A. formal measurement.
    B. evaluation
    C. validity
    D. reliability

3. Which of the following test item formats are typically easiest to construct but hardest to reliably score?

    A. Multiple choice.
    B. Matching
    C. Interpretive exercises.
    D. Essay

4. Making decisions about student performance based on a comparison to his or her peers best describes:

    A. performance evaluation.
    B. criterion-referenced evaluation.
    C. norm-referenced evaluation.
    D. reliable evaluation.

5. Giving a test or quiz for the purpose of gathering information and providing feedback, but not for assigning grades is called:

    A. informal evaluation.
    B. norm-referenced evaluation.
    C. formative evaluation.
    D. summative evaluation.

6. Which of the following is NOT considered to be an acceptable evaluation method for performance assessment?

    A. Multiple choice questions based on content objectives.
    B. Note-taking based on specific criteria.
    C. Checklists containing written descriptions of acceptable performance.
    D. Rating scales using explicit descriptors.

7. Which of the following is NOT a criticism of traditional assessment?

    A. They have low reliability.
    B. They tend to focus on low-level outcomes.
    C. They measure only outcomes, ignoring processes.
    D. They fail to tap learners' ability to apply knowledge.

8. Which of the following is the best example of an informal measurement?

    A. You give your students a pop quiz covering the multiplication of fractions.
    B. You devise a rating scale to assess each part of an in-class presentation by your students.
    C. You devise a set of criteria for scoring an essay test. You use the criteria in scoring your students' essays.
    D. You call on each student in your class in a question and answer session. And they're all able to answer.

9. Tests designed to measure and communicate how much students have learned in different content areas are called:

    A. performance tests.
    B. achievement tests.
    C. diagnostic tests.
    D. Aptitude tests.

10. Tests designed to predict potential for future learning and measure general abilities developed over long periods of time are called:

    A. Performance tests.
    B. Achievement tests.
    C. Personality tests.
    D. Aptitude tests.