

2-6 Examination/Introductions:

Is one method of evaluation of learners, which may indicate, assessment of knowledge, attitudes, skills, aptitudes, and all of them convey with the changes in learners' behavior. So the examination as one tool of evaluation aims to reflect that the goals of learning have been achieved. The learning experiences should be suitable to the subject's content and the measurement instruments.

Thus, the end-product of all education is the modification of the learners' behavior; therefore the education or the prepared curriculum (the planned experiences that are offered to the learners under the guidance of the school or teacher) is deliberate attempt by one or more persons to affect changes in behavior. (Behavior means any thing that a human being thinks or feels or does). The most important items here is how the teachers select the learning experiences, which emphasizes on an organized attempt to affect changes in behavior by presenting certain experiences, it implies conscious plan to instill some behavior patterns, and inhibit others. In short it is concerned with learning and the planning for learning, through the following:

- a- Learning is active process in which the learners must be involved.
- b- Learning is considerably affected by individual goals, values and motives.
- c- Immediate reinforcement promotes learning as a cognitive feedback.
- d- The wider the range of experiences presented to the learners, the more likely the generalization and discrimination are to occur.
- e- Behavior is the function of the learner's perception.

- f- Group atmosphere affects both learning product and satisfaction.
- g- Individual differences affect learning. Such differences are both biogenic and socio-cultural.
- h- All learning is multiple .though the focus may be on one particular (desired) outcome, so other learning takes place simultaneously.

Evaluation and learning:

Measurement and other kinds of assessment obviously play important role in the motivation of learning and in learning itself “where schools, teachers, and learners are rated to some extent by the test results, the nature of the test largely determines the quality of the schooling or Institutes process ...what teachers teach and how they teach , what learners learn ...and how they learn ". In addition the type of examination strongly influences the type of study procedure used by students preparing for it.

2-6-1 Examination:

Examination is a measuring instrument, one datum to be used in evaluation. Consideration of examination raises the question of standards. This term is commonly used in at least two different senses. "Examiner's' standards" are usually their estimates of what candidates should do and "maintenance of standards" usually means making the examination difficult, or not let too many candidates pass. (Can give different marks to the same paper at different times), no agreement among examiner or standards of individual paper. So the evaluation phase is concerned with actual out comes, not intended out comes or standards. Though all experts in measurement agree that objectives must be clarified and clearly stated before any assessment is attempted, one drawback to almost every subject syllabus or course,

and these are what most examiners based on (it is impossible to do any assessment without knowing what is to be assessed). This means that at all times goals and objectives determine assessment. If you need to test cognitive behavior in a single subject area (to assess knowledge alone) it is necessary to specify at least eighteen groups of objectives, two groups each for specific facts, terminologies, conventions, trends and sequences, classification and categories, criteria, methodology, principles, generalization, theories and structure. If comprehension is to be assessed, the three classes of translation, interpretation, and exploration can each be considered in relation to the nine knowledge sub-classes above. Application, analysis, and synthesis can each be applied to previously comprehended knowledge, while evaluation requires combination of necessary processes from any one or more of the five categories. Adequate assessment is possible only if specific objectives are detailed with reference to all these categories of particular subject (*D.K.Wheeler,1983: 81,281*).

2-6-2 Instruments of Evaluation:

There are variety of tools for assessment and evaluation of changes in behavior, based on its function. These measurement instruments are; intended tests, aim to examine intellectual abilities (intelligence tests), Psychometric tests (personality tests), social test (sociometric test), and academic performance tests (standardized tests or achievement tests). The researcher used the achievement tests, as the major technique used to assess academic promotion for the universities' students which is the basic topic of this research, and reflects the importance of examinations or tests. The essential functions of examination are;

1) - Make discrimination between students, (students with the high level of performance, from those within low level of performance).

2) - make categories for students according to their levels of performance, and this helps in;

a)- Distributing students in different specialization and schools or departments.

b) – Helping in assessing or evaluating students, which is necessary for evaluators to make judgments about promotion of the students in educational process or withdrawal the students.

c) - It is essential method in assessing or evaluating students' abilities and skills. There are two approaches in evaluating students' abilities:

2-6-2-1 Norm-Referenced Tests (NRT):

Norm-referenced test is a test designed to facilitate a comparison of an individual's performance with that of norm group. The norm group is some people refer to as standardized test. Is the national sample of examinees, but it may also be either a local population or sample of local population. (When your percentile rank is 64, you know that you scored higher than 64% of those in the norm group). These are:

2-6-2-2 Criterion-Referenced Tests (CRT):

Tests designed to measure the extent to which examinees have met standards performance. Standards performance are established by expert judgment and may be less than 100% correct, depending on the trait being measured. For example, on a typing test, we might require only 95%accuracy .Of course, performance standards are not always expressed as a percentage. We can use descriptive labels such as "expert", "novice", etc. is the criterion (that is performance standards). Notice that the interpretations of an examinee's test score is independent of how other student's performe. Being higher or lower

than others who take the same criterion-referenced test is irrelevant to the interpretation.

Those who build norm-referenced tests approach the task differently than those who build criterion-referenced tests. NRTs are intentionally built to be of medium difficulty –specially, items that are passed by only about 50% of examinees in tryouts are favored in the selection of items for the final version of the tests, because it facilitates the comparison of an individual with a group. We can delete the items of content is either very easy or very difficult in try-out, of the items. In contrast, when building a CRT, item difficulty is of little concern. Expert judgment is used to determine the desired level of performance and how to test for it- these judgments are influenced very little by how "difficult" a task is. Certainly, this approach is often appropriate. For instance, would you want the professors at a nursing school to drop the item on measuring the correct amount of medication before injecting it simply because it's too difficult for their students? Of course not; we'd want them to include the item regardless of its difficulty level.

2-6-3 Types of Examinations (tests):

All examinations or tests aim to measure the optimum of performance. When we measure achievement, aptitude, and intelligence we want examinees to show us their best- to perform at their optimum.

2- 6-3-1 Achievement test:

Measures knowledge and skills people have acquired. When we use achievement tests we designed it to measure the objectives of instruction, subject, they defined as valid when tests measured what should be measured.

2-6-3-2 Aptitude test:

It is designed to predict some specific type of achievement. An example is the College Board's Scholastic Aptitude Test (SAT), which is designed to predict success in college.

2-6-3-3 An intelligence test:

It is designed to predict achievement in general not any one specific type. The most popular intelligence test is cultural loaded and measure knowledge and skills that can be acquired with instruction. All people are exposed to such information and the more intelligent people are more likely to retain it. Efforts have been made to develop culture free intelligence test. Typically efforts in this area have concentrated on non verbal task, of the high cultural load of language usage. This test is usually less predictive of subsequent achievement than the more traditional intelligence test, because achievement is accomplished in cultural context (***Mildred L., 2002:71-73***) .

There are certain types of tests:

- a) Written tests (one type is essay test)
- b) Oral tests. Oral objective tests (concern with multiple of achievement tests, verbal, and practical tests).

2-6-3-4 Written Tests:

One form is an essay tests. This type of tests is used in all phases of education, especially at higher levels. The test is formulated in few questions which are not less than four questions, and not more than eight or ten questions. The questions verbs are started usually with the following; Explain, discuss, compare, analyses, criticize, ect. The answer may be in lines or one or more pages, and this depends on the nature of the subject. There is a sort of freedom for students to

reflect their own point of views that may adhering their abilities and skills. This freedom's ratio makes correction complex and may increases the subjectivity. It is criticized, but it is still being used in the university level, and this may due to the easiness in constructing questions, or may be the university staff members have not been trained in constructing objective tests (*Zakria Elsharief .elt, 1996:158*).

The advantages of the Essay Tests:

- a- Easy in preparing and constructing questions.
- b- A comprehensive instrument, that measures all levels of cognitive domain, start with understanding, and finished with evaluation, need high skills to be well designed, if teachers consider issues of the time for each question, and the length of the answer (how many words, or how many pages).
- c- Expand the opportunities for students to demonstrate their abilities, and creativity in analysis, synthesis, and criticize of subject's elements. This will help evaluators to distinguish between the students who understood the subject, and those who learnt it by heart only memorized it. But today the methodology of the briefed notes was distributed, and motivates students to deal more with it, because it contains basic information that is necessary for success (*Monera Elsaiede, 1991:236*).

The Disadvantages of the Essay tests:

- a- It is not valid tests because it is difficult to cover all subjects' content, but based on selected areas.
- b- Impact of personal factors, bias and subjectivity in correction.

- c- No stability of test's results, there is no specific standard for correction, if the paper is corrected by more than one teacher and even by the same teacher.
- d- External factors have affect on the teacher, e.g.; hand writing, organization of paper, language components, number of papers, psychological state of teacher, and environmental factors.
- e- Need a long time in correction, add to this a huge number of papers, so it needs efforts to continue in correction which may be divided in different sessions (variation in correction standards within each session).
- f- Ambiguous questions may discriminate between students according to their understanding of the questions , and may reduce the motivation of the students in completing answers, which may raise level of worried and anxiety , and it seems as an aspect of psychological impact of assessment .

2-6-3-5 Objective Tests:

It is a recent type of testing which came out from the process of the development of evaluation. It is an information source for evaluation, and it concentrates on the formative type of evaluation (continuous assessment). It seem as essential in formative evaluation, because it reflects an accurate picture of students' performance, and it offers possibility to assess the whole elements in subject matter, with high ability to assess and evaluate students in all levels of cognitive domain objectively.

These types of objective tests have advantages and disadvantages on both teachers and students.

❖ **The Advantages of Objective Tests:**

a)-On Teachers:

-Easy in correction, if it is prepared well and adequate, the same test can be used for long time. It is easy for teachers who are constructing objective tests to create questions' bank of objectives tests.

b) On the Students:

- May motivate students when the teacher diagnose his/her weakness, and help in over come weakness. It is a valid method to assess students' knowledge, and abilities. But it must be characterized by validity, reliability, objectivity ... and so on.

❖ **The Disadvantages of Objective tests:**

a)-On Teachers:

-It put a surcharge (boredom and exhaustion) on teachers, due to continuous building of testing, correcting it, and to make analysis for results, and this needs time and efforts.

b)- On Students:

-May put students on continuous stress, due to continuous assessment during the year.

(Muna. M, 1997:36).

2-6-3-5-1 Types of Objective Tests:

Include many types; the following are some of them:

1)-Short –Answer Tests:

This type of test is used to measure the lower level of cognitive domain, which is concerned with understanding and comprehension levels. It contains the following types:

2)-Free –Answer test:

The answer is by a word, sentence, or number.

3)-Completion –Test:

The answer is by filling in the blank spaces by adding a word, or phrase that is missing. This type of test is measuring memorizing. Both short answer and completion tests, are used in measuring understanding, comprehension, and description, and Cover major parts of subject content, rather than the essay test

4) - Multiple-Choices Test:

It is the most difficult type of objective tests used to assess academic performance. It is used to assess lower levels of cognitive capabilities, and not recommended to assess higher levels of knowledge, because it needs a professional expert. The structure of it is, a sentence with specific trunk or stem, followed by three, four or five options only one is correct, or may use the technique of presenting all options are correct and the answer is the most correct answer. This type of test measures both lower level and high level of cognitive domain for individual.

The principles for constructing multiple –choices questions:

There are certain bases for constructing multiple-choices questions those are:

- a) The question must present specific and clear problem with its introductory stem, with complete and understandable sentence or statement.
- b) The introductory stem should be combined between all others choices, with suitable length.
- c) Similarity between all options that contain correct answer among fault answers.
- d) Avoidance of unintended insinuation in introductory stem, or among choices and this appears in generalization or

specification terminology, which may serve as an indicator to the correct answer.

- e) The structure and the length of the correct answer should be similar to the rest of choices.
- f) Avoid to use clauses or phrases like; all mentioned above statements are correct, or all mentioned above statements are wrong, or; a- and c- are correct (*Abdelrahman Ahmed, 2002:31*).

The advantages of Multiple –choices Test include the follows:

- 1/ it is Characterized by high level of validity and reliability.
- 2/ it is a comprehensive in covering all items of subject content.
- 3/ it is free from bias or subjectivity.
- 4/ it is easy in statistical calculation.

The Disadvantages of Multiple-Choices Test:

- a- It is expensive in the typing materials, number of papers, and so on.
- b- It is complicated in building, or designing its structure, it needs experts with scientific training, long experiences, and specialization.
- c- It is offers high possibilities in cheating.

5)- True and False Test:

This type is designed in away that the student answers with assigning the question or the statement by true, yes or in fault can select false answer or no in special place for answer (between practice) , mainly this type deals with the facts or information that is accurately true or false in subject matter. The problem of it is the guessing problem, the probability of guessing correct answer may reach to 50%, thus will indicate invalid test in assessment of student's performance. The educational psychologists try to reduce

the probability of guessing the correct answer through student's justification of his/her answer:

- a) If answer true or yes justify why? If answer is fault or no, so write the correct answer.
- b) Applied guessing quotient and this is through;
- d- $\text{Number of correct answers} - \text{number of wrong answers} = \text{student's mark.}$
- c) Write clear statement without specialization or generalization of it, with specifying in one idea.
- d) Avoid negative phrases, because they may hesitate student's attention
- e) All statements or questions must be equal in phrases and so on. (*Robert Thorndike & Hagen, 1989:209*).

all these types of objectives questions are easy to correct, and covered the majority of the parts of subject's content, but the disadvantage of it is to help the learners on memorizing, it can't measure the high levels of cognitive domain like; analysis, synthesis, and evaluation.

7)- Oral Examination:

This type is recently used in some colleges at the end of studying years and before graduation. It can be used as a comprehensive oral exam, aim to expose students' personality and abilities; also it is used to certify high scientific degrees. It is characterized by less in costing, no way for cheating or guessing in answers. The disadvantages of oral examination are; high level of subjectivity that is based on personal influences (subjectivity in assessment), takes time and efforts because it is individualized test, also it is injustice in distribution of questions for students in rating easy or complex questions. Less reliability and validity, because it depends

on fluently of students' language, which is not an objective tool to evaluate students, and may reflect some sort of biasing (*Zakria Elsharief . et al, 1996: 165*).

8) - Practical Tests: It is a type of testing, being used mainly in higher education and is assumed as primary part of formative evaluation (continuous assessment). It measures the higher levels of cognitive domain mainly, and it involves both written and oral tests, and measures the three domains of individual development: cognitive domain, psychomotor domain, and affective domain. It needs a large number of university teachers, compared with the increasing university students' numbers. It is used in expose the level of practical performance in subjects matter. It is necessary to some extent in technical colleges. This test is characterized by being closer to criterion-referenced test, because it deals with revealing the optimum level of student's performance, thus will discover how students benefit from theoretical knowledge and how facilitate it. It is very adequate test, so the products (results) are accurately valid and reliable. The disadvantages of it are; consumes times and efforts for both teacher and student. It shapes some sort of subjectivity if evaluation procedure done by only one person, and this will reduce its objectivity.

2-6-3-6 Test's Grades:

The most important issue after structuring a test is to plan how to distribute the marks across the test's questions. The correction and grade system is initially for evaluation process. The tests at universities are affected by certain factors which may reduce or increase the students' grades, these factors are; environmental factors, social factors, individual differences, and all are distributed normally according to Gauss curve (normal distribution curve) ,

which are uncontrollable factors, it is defined as statistical-probability factors. Other factors related to educational instruction are; nature of the subject, teaching methods, time consuming, practical training, experiments, which are controllable factors. The teacher is responsible for constructing grades 'or marks' stair (set of steps for grading). In which bases the teacher should decide the marks for each question. This helps the teacher to be more reliable in correction process. The marks' stair includes; the distribution of the students' results (normal curve distribution), concerns with factors like; difficulty of subject, absence of experiments, limitation of time, number of the students, and a few of references, specify the minimum level for passing. The final result is a normally distribution according to Gauss Curve, both controllable and out of control factors will influence the result, the solution is through standardization of results. Today most of the universities are using the computer in standardization of result to be more acceptable and near to the normal distribution.

2-6-3-7 Statistical Analysis for Test's Questions:

It is a procedure in analysis of each question to find out easiness and difficulty coefficient, to discriminate index of questions, and make judgment on excluding it or not.

Easiness coefficient:

Is a method used to calculate the percentage of students who answered the questions correctly, and is divided by total number of the students. The highest easiness coefficient is usually =1.

$$\text{Easiness Coefficient} = \frac{\text{Number of students answered questions correctly}}{\text{Total number of the students answered question}}$$

Difficulty Coefficient:

Find out the percentage of the students who answered the questions wrongly, and divide by the total number of the students who answered the question. The highest difficulty coefficient is usually =1.

$$\text{Difficulty Coefficient} = \frac{\text{Number of students answered questions wrongly}}{\text{Number of the students who tried to answer question}}$$

When we add the easiness coefficient to difficulty coefficient it equals one. Easiness coefficient + Difficulty coefficient = 1.

Discrimination Index:

The easiness coefficient is one indicator used to assign the ratio of student who answered correctly, then passed the question, while difficulty coefficient used as indicator to account percentage of students who answered faulty. But both of them do not reflect the interpretation about the nature of the students who passed or failed. The discrimination index used as indicator to discriminate among students, those who are, excellent, good, average, and weak or failure students, through the question it self (*Safout Farag ,2000:135-137*).