University of Echahid Hamma Lakhdar - El Oued
Faculty of Social and Human Sciences
Department of Psychology and Educational Sciences

Educational Assessment Course

A Publication for Third-Year Educational Sciences Students Specializing in Educational Psychology and Guidance and Counseling

Authored by: Dr. Abdellatif Guenoua

Academic Year: 2023-2024

Educational Assessment Course

Educational Assessment Course
A Course Designed for Third-Year Educational Sciences Students
Specializing in Educational Psychology during the First Semester and Specializing in Guidance and Counseling during the Second Semester

Educational assessment is one of the fundamental components of education and pedagogy. This article aims to explore its conceptual framework, historical evolution, various types, governing theories and models. Through this exploration, we will examine educational assessments within the Algerian educational system, drawing insights from implemented curricula. Subsequently, we will analyze assessment tools and strategies such as achievement tests, performance assessments, observation grids, interviews, and portfolios. Additionally, we will address self-assessment and peer assessment, which align with the evolving concept of assessment. Given the technological advancements in the field of information and communication technology, we will introduce the concept of electronic assessment mechanisms and question banks. Finally, we will discuss some of the challenges encountered in educational assessment.

Table of Contents

The Concept of Educational Assessment and Its Development Throughout History

1. Historical Overview of the Development of Educational Assessment

Educational assessment, as a research field and an independent academic discipline, emerged relatively recently, in the second half of the previous century. However, its practice dates back to ancient times. Early humans evaluated various characteristics, comparing themselves or others based on attributes such as strength, height, usefulness, or suitability for specific professions. Likewise, they were interested in gauging their progress in their work or the profits they earned in their trade, assessing how close they were to achieving their goals.

In the field of education, the Chinese were the first to use a tool for assessment, specifically oral examinations, to evaluate progress and eligibility for jobs (Mahmoud, 2004, p. 15). With the advent of Islamic civilization, oral questioning was employed to assess students' understanding of knowledge and their ability to memorize the Quran, Sunnah of the Prophet, and other significant texts that emerged.

In the late 19th century, the movement of psychological measurement and the emergence of psychological testing laid the groundwork for the development of mental tests, which were primarily implemented in schools and administered to students. This marked the establishment of a connection between psychological measurement and educational assessment.

In 1864, George Fisher designed the first written achievement test, comprising several graded measures ranging from easy to difficult, which were used to evaluate specific prescribed school subjects. These tests were referred to as "*Scale-Book*" (Allam, 2002, p. 303).

Furthermore, the first standardized objective test was created in the United States in 1895 by Rice and administered to a large cohort of fourth- and eighth-grade students. This initiative marked the inception of standardized

achievement tests, which gradually began to emerge thereafter (Alam, 2002, p. 305).

Thorn-dike's research emerged, emphasizing the necessity of integrating tests into the field of education, with a particular focus on positioning assessment as an essential component of the teaching process.

These tests began to proliferate, leading to the development of test batteries, such as the Stanford Achievement Test Battery for elementary school in 1923 and the Achievement in Writing and Arithmetic (AWA) battery for secondary education in 1925 (Qureshi, 2002, p. 63).

Tyler's model of assessment emerged, expanding its scope beyond merely evaluating students. It broadened the concept of assessment to encompass the entire educational program, focusing on the extent to which its objectives were achieved.

In addition to moving away from solely emphasizing students' cognitive aspects, Tyler's model highlighted the importance of emotional and psychomotor dimensions. This approach also marked a shift from relying exclusively on written or oral tests to embracing a variety of assessment methods.

The continuous evolution of educational assessment underscored its significance within the educational process. Assessment became a fundamental subject taught to students in teacher training institutes and colleges of education. Over time, it developed into an independent research field, fostering knowledge production and establishing itself as an academic discipline, as seen in Algerian universities.

Contemporary trends in assessment have reduced the reliance on standardized tests, placing greater emphasis on essay questions, observation, and evaluating students' achievements in ways closely aligned with real-life contexts. The American Association for the Advancement of Science advocated for the use of authentic assessment, which provides more reliable insights into student learning. Studies have demonstrated improved educational outcomes with this approach, commonly referred to as authentic assessment (Shehata, 2013, p. 323).

Through tracking this development, we can record the following observations:

Through tracking this development, we can record the following observations:

- A shift from assessing students' memorization, recitation, and comprehension at lower levels of cognitive objectives to higher-level goals involving complex thinking, such as analysis, synthesis, critique, and problem-solving.
- A shift from evaluating students solely in the cognitive domain, which previously received the most attention, to assessing the affective and psychomotor domains.
- A shift from using a single method, represented by achievement tests, to employing multiple means and tools including observation and performance assessment.
- A shift towards authentic assessment by observing students' behavior in real-life situations, rather than evaluating them in artificial scenarios that may be disconnected from their realistic practices.
- A shift from assessing students to assessing all aspects related to the educational process, including goals, content, methods, and resources.

2. Definition of Educational Assessment:

The concept of educational assessment originated from the linguistic concept, which has two meanings: the first meaning is to rectify, modify, or improve, and the second meaning is to assign value to something and assess its worth. Al-Fayruzabadi (2008) stated in *Al-Qamus Al-Muhit* that the term refers to the process of eliminating deviations, evaluating the merchandise, assessing its fairness, and determining its price.

In the field of education, researchers have provided various definitions of educational assessment, including:

- **Bloom et al. (1971)** define it as a judgment of the value of ideas, actions, solutions, methods, and materials for a specific purpose. It involves using criteria, levels, or standards to assess the adequacy,

accuracy, and effectiveness of things, and assessment can be quantitative or qualitative (Khattab, 2001, p. 5).

- **Gronland (1971)** defines it as an organized process aimed at determining the extent to which students achieve educational objectives (Mikhail, 2015, p. 37).

- **Stake and Screven (Alam, 2015, p. 19)** define it as the process of obtaining information and making judgments. Screven adds that it also involves making decisions based on the gathered information.
- **Alam (2015)** defines it as a systematic process involving the collection of objective data and reliable information through various measurement tools. This process is guided by a set of expected levels or defined objectives, aiming to produce quantitative estimates and qualitative evidence. These serve as the basis for making informed judgments or appropriate decisions regarding students and the educational process. The ultimate goal is to improve performance quality and enhance proficiency, aiding in the achievement of these levels or objectives.

Based on the definitions, we find that they emphasize key aspects of the concept of educational assessment, with variations in their presentation according to each researcher's perspective. These key aspects include:

- It is an organized and planned process that follows specific procedures and employs reliable tools.
- The purpose of assessment is to collect data and information for quantitative or qualitative estimation.
- Judgments are made based on criteria, levels, or specific objectives.
- Decisions are made for improvement, adjustment, and enhancing proficiency, which represents the ultimate goal of educational assessment—a concept that cannot be fully understood without achieving this aim.

3. Assessment in Relation to Measurement and Evaluation:

The relationship between assessment, measurement, and evaluation can be better understood by examining their definitions. When defining educational

assessment, we notice that it intersects with two related terms: measurement and evaluation. Before clarifying the relationship among them, we define measurement as the process of quantifying and estimating characteristics or attributes. Alam (2002) defines measurement as the assignment of a set of numbers or symbols corresponding to the characteristics or attributes of individuals, following well-defined rules. This indicates that educational measurement focuses on quantifying the traits or attributes of individuals, as it is impossible to measure individuals themselves directly; instead, we measure their traits or characteristics.

Measurement, in this context, is encompassed within the definition of assessment. Assessment is broader and more comprehensive than the measurement process, which serves as an essential component of any evaluative procedure.

On the other hand, evaluation refers to the process of gathering, organizing, and interpreting information and data to support decision-making regarding students and to guide them based on their abilities and inclinations. Information is collected through performance tests and observation. However, this process does not include diagnosis, treatment, or prevention, which are specific to the assessment process (Khalil, 2011, p. 5). From this definition, it becomes clear that the term evaluation lies between measurement and assessment. It is broader and more inclusive than measurement, which focuses on processes and tools, but narrower than assessment, which incorporates modification, improvement, and intervention. Therefore, assessment is the overarching and more comprehensive term.

4- Objectives of Educational Assessment:

Educational assessment serves several general and specific objectives, some of which include:

- Evaluating the extent to which goals are achieved, whether at the level of a specific lesson or session or broadly across a program or curriculum.

- To gather qualitative and quantitative information based on standards and levels about the actual performance in educational activities, whether from the perspective of learners, teachers, or educational administrators.
- To identify strengths and weaknesses to develop remedial plans or implement changes in curricula and programs.
- To identify students' readiness, abilities, and challenges.
- To determine the causes of success or failure.
- To support students in achieving academic alignment and enhancing their performance.
- To make decisions and judgments about students, such as determining their success, failure, or directing them to specific fields or specializations.

5- Functions of Assessment:

Educational assessment has multiple functions, which can be summarized into three main functions:

5-1 Guidance:

Assessment is conducted when we want to guide students towards specific fields of study or specializations. The guidance process is based on the abilities, potential, acquired knowledge, and qualifications of students. Therefore, the assessment process precedes the guidance process. Additionally, assessment allows us to have an evaluative perspective on the obstacles that may hinder students' progress. The assessment process may cover various aspects of the student and not be limited to academic achievement assessment alone.

5-2 Adjustment:

The adjustment function is a crucial aspect of the comprehensive concept of assessment. It involves diagnosing learners' errors, categorizing and analyzing them, and developing a remedial plan. This can be achieved through changing methods or approaches, utilizing different means, and providing support that helps students achieve educational goals and acquire the planned competencies. The adjustment function is performed by formative assessment,

which takes place during the teaching and learning process, providing feedback that allows for modification, improvement, and error correction.

5-3 Certification Function:

This function is evident at the end of the learning process by awarding a certificate that verifies the learner has acquired the required competencies, whether for a class, academic year, or educational stage. Through this function, an endorsement or certification is issued, affirming the student's success and qualification to advance to a certain level or obtain a specific degree. This function is commonly referred to as summative assessment.

6- Steps of Assessment:

Assessment is an organized process that goes through specific steps, which can be summarized into four main steps:

6-1 Identification of Purpose:

Identifying the purpose and objective of the assessment, such as whether it is diagnostic, summative, or formative, and whether the goal is guidance, certification, or adjustment. Also, determining the scope, content, and competencies that will be the subject of assessment.

6-2 Collection of Data:

Gathering the necessary data and information through reliable and organized tools and means, such as tests in their various forms, observation, or any other valid measurement methods. Then, arranging and classifying this data, whether quantitative or qualitative, and conducting the process of analysis.

6-3 Issuance of Judgment:

This involves making decisions based on objective data rather than preconceived judgments or subjective opinions. It requires comparing the data to external criteria, predefined standards, or specific levels.

6-4 Decision Making:

After issuing a judgment based on the collected data provided by the measurement, we make appropriate decisions regarding the educational path, such as advancement to a higher level, repetition, redirection to specific branches, or reorientation. Additionally, there may be adjustment decisions

through a remedial plan for errors or decisions related to teachers and educational administration, particularly in the pedagogical aspect.

7- Areas of Assessment:

7-1 Learner:

This is the most important area on which assessment focuses, as the learner is the center of the educational process. Assessment aims to uncover their readiness and abilities for appropriate utilization, identify the difficulties and obstacles impeding their progress, evaluate the levels they reach periodically, or permanently, for the purpose of adjustment, improvement, achieving harmony, and academic success, and inform them about their progress towards their goals.

7-2 Teacher:

Educational assessment includes the teacher, who is responsible for the teaching process. They are the guide, assistant, and supervisor of their students' learning. In order to ensure their effective performance, the teacher should be included in the assessment process regarding their teaching competencies. This includes monitoring to identify shortcomings, difficulties, and problems. This information is studied to intervene and improve through suitable professional development.

7-3 Curriculum Content:

Curriculum content is constantly subject to continuous renewal due to significant and rapid developments in technical, social, economic, and political aspects. This necessitates the use of assessment for this content as time progresses or when there is a general need or specific requirements for students.

7-4 Teaching Methods:

The success of the educational process is linked to the activities that take place within the classroom and the methods employed by the teacher in instruction. This calls for the assessment of these methods and the study of their effectiveness and suitability for the learners. The appropriateness of a teaching method is determined by the results it achieves. Therefore, teaching methods are constantly evolving, being modified, and improved.

7-5 Teaching Materials:

To facilitate the educational process, various instructional materials are used, ranging from chalkboards and pens to laboratories and maps. The use of multimedia resources is also evolving continuously. These materials are subject to assessment to explore the utilization of modern technologies in education. This has led to the emergence of e-learning, question banks, and electronic assessment, and, therefore, the assessment of teaching materials has become necessary.

7-6 Assessment of Assessment:

Assessment is an integral part of the educational process, and it employs various tools such as written or oral tests, observation, performance assessments, and more. Therefore, the assessment of assessment itself is considered a domain within the field of assessment. It involves evaluating the effectiveness and suitability of these tools, their construction, development, and the discovery of new assessment tools, along with assessing the competencies of teachers in this regard.

Types of Educational Assessment

There are several classifications of assessment based on specific characteristics or aspects that allow for dividing educational assessment into various types. Among these classifications are the following:

1- According to the time duration

Educational assessment can be categorized based on the progression of the educational process, which unfolds in three fundamental steps or stages. These stages necessitate three distinct types of assessment:

- Diagnostic Assessment: Conducted before or at the beginning of the lesson.
- Formative or Constructive Assessment: Implemented during the stages of the lesson.
- Summative, Final, or Achievement Assessment: Administered at the end of the lesson or stage.

This classification underscores the most significant types of assessment.

1-1 Diagnostic Assessment: (Sareer & Khaldi, Eds., 95)

It is the determination of students' performance at a certain stage (especially at the beginning of the academic year) to organize reinforcement programs based on that, such as university entrance exams. Diagnostic assessment identifies students' grades, areas of failure, and reasons such as hearing or visual impairment. Its goal is to obtain data and information that indicate the extent of students' mastery of previous acquisitions (abilities, skills, knowledge) upon which the teaching of new data is based. Diagnostic assessment prepares for and assesses prior achievements.

1-2 Formative Assessment: (Sareer & Khaldi, Eds., 98)

Scriven stated that it is the continuous assessment of students throughout their learning process (especially at the end of a unit) to enable teachers and curriculum developers to correct the course of learning before its completion. Its tools include observing students, asking them questions, giving them

classroom and homework exercises, and tests. It is a continuous process followed throughout the lesson period, serving as an evaluative test informing about the conditions and situations of teaching and learning at that time. It also helps determine the extent to which the learner is keeping up with the lesson and the difficulties that may arise, enabling the teacher to quickly address weaknesses, reinforce strengths, and correct difficulties, thus achieving the goal.

When teachers use data to make decisions about the next steps for the student or group of students, it means forming and building practices; thus, this is formative assessment. However, when data is collected at specific and different intervals and used to indicate what students have achieved compared to that data, it presents a summary of students' progress; thus, this is called cumulative assessment. (Biggs, 2014, p. 7).

1-3 Summative Assessment: (Sareer & Khaldi, Eds., 102)

It is conducted at the end of a lesson, semester, or application of a specific curriculum to be utilized in awarding promotion certificates to students or in curriculum development upon its revision.

2.Types of Assessment According to Purpose: (National Assessment Committee, 2004, p. 25).

Assessment can be classified according to its purpose into several types, including:

- Diagnostic Assessment.
- Assessment for the purpose of adjusting plans and curricula.
- Assessment for student classification purposes.
- Assessment for selection or admission purposes.
- Assessment for prediction and identification of readiness.
- Assessment for checking learning and confirming the quality of outcomes.
- Assessment for confirming the effort exerted.

3. Criterion-Referenced Assessment and Norm-Referenced Assessment:

When interpreting assessment results, two approaches can be used:

1. Criterion-Referenced Assessment: Results are compared to a specific external criterion (e.g., achieving a score of 10 to pass an exam).
2. Norm-Referenced Assessment: Results are interpreted relative to the group to which the individual belongs (e.g., ranking candidates in a doctoral competition regardless of their actual scores).

These two types can also be combined. For example, in teacher recruitment competitions, an external criterion (e.g., achieving a score of 10) may be applied first, followed by ranking those who meet this criterion to select the required number of candidates.

4.Cognitive Assessment and Performance Assessment:

Cognitive assessment is based on the extent to which cognitive objectives are achieved, according to Bloom's taxonomy, for example, remembering and understanding as lower levels and analysis, synthesis, and evaluation as higher levels. This means looking at what the learner knows after being presented with a specific course or part of a particular subject.

Performance assessment, on the other hand, evaluates what an individual can do or apply, relating to their external reality. It assesses the skill aspect through tasks performed by the student with a degree of proficiency based on what they have learned in the lesson.

5.Quantitative Assessment / Qualitative Assessment:

Quantitative assessment relies on collecting numerical data about students' achievement, while qualitative assessment relies on collecting descriptive data about students' achievement and competencies acquired. This is often done through observation, especially when we want to understand some characteristics, inclinations, attitudes, and preferences of learners, beyond numerical measures. These two types of assessment complement each other.

6- Macroscopic Holistic Assessment and Microscopic Partial Assessment (Khattab, 2001, p. 13):

Macroscopic assessment involves evaluating the outputs of systems as a whole in relation to the objectives of the system's overall public policy, such as ensuring an educational goal from general objectives, like providing the community with professional competencies. On the other hand, microscopic assessment addresses aspects of the institution or the object being evaluated without reaching a general evaluative perspective. It can be an assessment of components individually or an assessment of dimensions or aspects. This is what is done in evaluating student work through tests.

7- Traditional Summative Assessment and Authentic Assessment:

Summative assessment, or classroom assessment, involves the use of achievement tests to evaluate learning outcomes and what has been taught, regardless of its connection to real-life situations or the problems students face. For example, teaching a student how to pray and testing them on it, without considering whether they perform it correctly and at the appropriate time. In contrast, authentic assessment immerses students in tasks that are meaningful and relevant to their lives, revolving around real-world problems they encounter. These tasks appear more as learning activities than traditional tests. In this form of assessment, students develop higher-order thinking skills and integrate a wide range of knowledge to make judgments, make decisions, or solve real-life problems they face (Shahata, p. 314).

8- Self-assessment and Peer Assessment:

Assessment can be categorized based on the evaluator conducting the assessment. Self-assessment is performed by the student or individual themselves to evaluate their progress, identify mistakes, and correct them. Peer assessment, on the other hand, is conducted by the teacher or institution, such as through midterm exams and official tests. Recently, peer assessment, where students evaluate each other, has emerged as a form of assessment.

9- Minimum-Standard Assessment and Mastery Level Assessment:

Minimum-standard assessment is designed to classify learners as either capable (successful) or incapable (unsuccessful) based solely on whether they meet the minimum standard, without considering variations in performance quality. For example, evaluating a person's ability to drive by determining whether they are capable (issuing a license) or incapable (withholding the license), without assessing the levels of skill demonstrated in this ability. In contrast, mastery level assessment categorizes a specific ability by referring to the possible levels of proficiency in the field. It applies mastery standards, and the minimum standards must be met. This involves evaluating the acquisition of competence while determining the level of proficiency through grades or feedback.

10. Continuous Assessment / One-time Assessment

Continuous assessment is an ongoing process that accompanies the learning journey, aiming to help students achieve mastery of the required skills and knowledge by providing feedback to guide and correct their educational path, thus supporting the continuation of the learning process (Al-Harthy, p. 18). This form of assessment overlaps with formative assessment. In contrast, a one-time assessment involves determining grades or making decisions based on an evaluation conducted at a specific point in time, typically at the end of a course or the beginning of the next. In this case, the focus is solely on what the learner is capable of doing at that moment, without considering prior performance. While assessment is often seen as separate from the learning process, it typically occurs at specific times and leads to decision-making. Continuous monitoring, however, integrates assessment into the course itself, contributing cumulatively to the final outcome. This process is supported by homework assignments and periodic knowledge tests or regular assessments that enhance learning. In contrast, one-time assessments focus on ensuring that an individual is capable of performing tasks, while continuous assessment is better suited to reflecting creativity and a broader range of skills.

Some Models of Educational Assessment (Taylor, Scriven, Stufflebeam, Stick, Hammond)

Research and study within the general framework that governs assessment lead us to propose models and theories that represent and summarize the assessment process, providing an image and reflection on its components and what assessment focuses on.

1. Definition of Assessment Model (Al-Shamari, 2013)

The term "model" linguistically means an example or representation of something, a form taken as an example of that thing to understand its condition. Specifically, an assessment model is a theoretical conception of the necessary steps to reach a specific decision regarding the object to be assessed and to report that to the stakeholders. It is a work plan simplified by the model's creator, who believes that following it is capable of producing effective assessment. In other words, the model is a conceptual framework for assessment that describes the creator's vision of the fundamental assessment concepts and the methods by which these concepts can be employed to reach justifiable judgments and recommendations.

Assessment models are used in two main forms: prescriptive models, which are the most common, consisting of a set of rules, descriptions, cautions, and guiding frameworks that define the nature of good assessment and how it should be conducted, and descriptive models, which are sets of statements and generalizations that describe, explain, or predict assessment activities.

Assessment models vary according to the researchers who have studied the subject of assessment. They have evolved significantly in response to calls for assessment reform, aiding in the development of assessment studies and the establishment of a common language that facilitates scientific communication and the exchange of expertise in this field.

It should be noted that researchers conducting assessment studies are not obliged to follow a specific model. They can combine multiple models that serve their work or create their own model that serves their assessment study. Given the existence of numerous assessment models, the different classifications of these models from the perspective of education specialists will be presented, along with an overview of some of these models.

2- Classification of Educational Assessment Models (Al-Muhaimid, Al-Siriy, Al-Ajlan, 2014):

There are several attempts to classify educational assessment models, which we will review as follows:

1. Bofam's Classification:

- Goal-Oriented Models (Objective Assessment Models): These models, including Taylor's model, view assessment as the process of determining the extent to which objectives are achieved.
- Judgmental Models: These models are further divided into two types:
  - Input-Focused Judgmental Models: Assessment is based on internal criteria, guided by the professional judgments of the assessor.
  - Output-Focused Judgmental Models: Assessment is based on external criteria, evaluating the outcomes of the assessed subject.
- Decision-Making Models: These models aim to provide information for decision-making purposes.

2. Wernimont and Sanders' Classification:

This classification is the most common and divides assessment models into the following categories:

- Goal-Directed Assessment Models: These models view assessment as the process of determining the extent to which objectives are achieved. Taylor's model is one of the prominent examples.
- Goal-Free Assessment Models: These models allow assessors to explore all program effects, whether intended (related to objectives) or unintended (unrelated to objectives). Scribner's model is an example.

- Administrative-Serving Assessment Models: These models aim to provide information about the program to decision-makers for program evaluation. Stufflebeam's model is an example.
- Natural Assessment Models and Participant-Oriented Assessment Models: These models describe the program as it is without attempting to control variables. Assessment is conducted naturally and addresses the issues and concerns of the stakeholders involved in the assessment. Examples include the Stick Model and Consumer-Oriented Assessment Models.
- Expert Judgment-Based Assessment Models: These models involve a team of experts issuing judgments regarding the worth or effectiveness of the subject being assessed.

3. Tyler Model (Al-Muhaimid, Al-Siriy, & Al-Ajlan, 2014):

Ralph S. Tyler (1902-1994) is among the pioneers who wrote in the field of curriculum in his book "Basic Principles of Curriculum Instruction." Taylor is considered the father of assessment, being the first to introduce organized assessment into education. His perspective on educational assessment, which he applied during the eight years he supervised in the 1930s, formed the basis of this model. His model (1949) contributed to driving the curriculum development process worldwide, serving as the guide for subsequent models.

Tyler's model is a goal-oriented assessment, belonging to objective models. Education is perceived as a process of behavior change through formulating operational goals that students perform, can be measured, and then assessing the extent to which these goals are achieved.

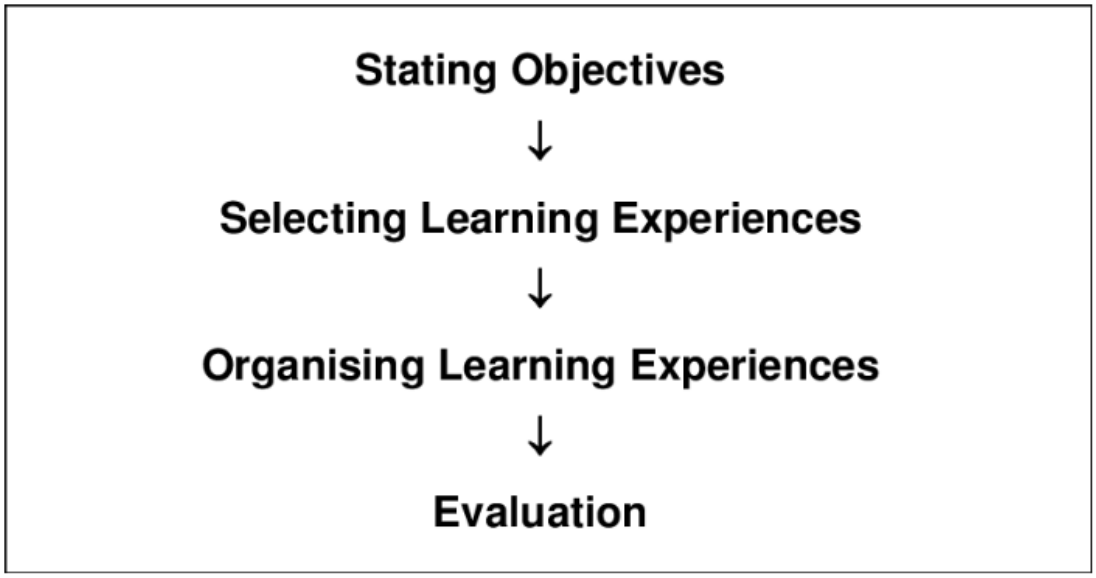
Taylor's model aims to link goals with action plans by making them behavioral goals; they reflect changes that occur in participants as a result of receiving sufficient instructions or gaining sufficient experience. This necessitates standardized tests for participants before and after exposure to these changes.

According to this model, assessment proceeds through a series of steps:

1. Identifying general (ultimate) and behavioral objectives to be achieved.
2. Classifying those objectives.

3. Translating each general objective into a set of operational (behavioral) objectives.
4. Selecting, organizing, and directing learning experiences.
5. Constructing and developing assessment methods and tools, or selecting them from existing tools.
6. Collecting data on student performance and comparing it with the formats of behavioral objectives.
7. Evaluating experiences, meaning reaching final decisions related to predetermined objectives.

Fig 1: Tyler's Evaluation Model



4. Scriven Model:

Also known as Goal-free evaluation (GFE), its concept was coined by Scriven to emphasize the need to inspect the actual effects of the program, whether intended or unintended. This can be achieved by developing an evaluation concept independent of the program's stated objectives, allowing employees to perform better without being aware of the program's official goals.

Scriven distinguished between two types of evaluation:

1.Formative evaluation involves assessing and improving an existing program by gathering necessary information from various sources.

2.Summative evaluation, which involves identifying the final outcomes of a completed program and collecting information to either continue, modify, or terminate the program. Additionally, there is a need to evaluate the program's impact, whether direct or indirect, intended or unintended. Scriven emphasized the importance of comparative evaluation, comparing the program with others to assist in selecting available alternatives. (Doumis, 2008, p. 123).

Scriven emphasizes not limiting evaluation solely to goal-based or goal-free evaluation but considering each as complementary to the other to ensure a comprehensive evaluation of the subject matter.

Scriven's evaluation is consumer-oriented, focusing on serving the consumer or beneficiary of the educational program. He views goal-free program evaluation as akin to evaluating a commodity or product by obtaining feedback from users and beneficiaries rather than the product producers. (Al-Shammari, 2013).

5-Stufflebeam Model (Abu Anab, 2000):

This model, also known as the "Context, Input, Process, and Product" (CIPP) Evaluation Model, is designed as a general framework to assist managers facing four different types of educational decisions related to inputs, content, processes, and outcomes. The model's name is derived from the initials of each element in English:

1. Context Evaluation:

This serves planning decisions by identifying needs that help define the objectives of the educational program.

2. Input Evaluation:

This serves construction decisions by determining the capabilities and resources of the entity entrusted with implementing the program. It involves reviewing alternative plans and evaluating them in terms of cost, benefit, duration, and potential obstacles to identify suitable plans.

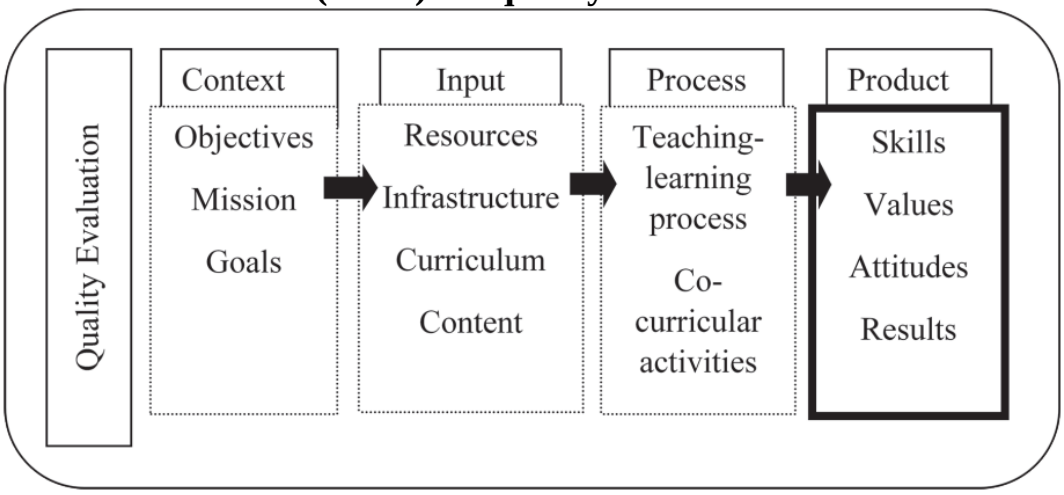
3. Process Evaluation:

This serves as a formative evaluation of the program by collecting data on its progress, implementation, encountered obstacles, and appropriate modifications for adjusting procedures. It provides feedback to decision-makers regarding program implementation.

4. Product Evaluation:

This serves as the summative evaluation of the program, aiming to determine the extent to which the program's purpose and objectives have been achieved. It links to content, inputs, and processes, serving decision review, and is used to determine program effectiveness. The following diagram illustrates the relationship between the four types of evaluation and their corresponding types of decisions:

Fig 2. Conceptual framework of implementation of Stufflebeam model (CIPP) for quality evaluation



Source: Shamsa et AL. (2018)

6. Stake's Model (Al-Shammari, 2013)

The Stake Model, as discussed by Al-Shammari (2013), approaches evaluation from a perspective fundamentally different from Taylor's. Stake criticizes Taylor's model for relying solely on objectives in the evaluation process. Stake argues that objectives may change during implementation, necessitating continuous communication between the evaluator and the

program beneficiary to discover and consider important program-related matters during evaluation.

Furthermore, Stake criticizes Taylor's evaluation for its narrow focus on specific aspects of the program (achieving objectives only) and its limited method of collecting evaluation information, which is restricted to standardized tests, tools, and experiments.

Stake believes that the purpose of evaluation is to help the beneficiary form a comprehensive understanding of the evaluation subject. Stake emphasizes the importance of avoiding issuing final judgments about the evaluation subject and instead advocates for collecting and presenting others' opinions and judgments, along with one's own, to the beneficiary in a suitable format.

Stake contributed to the development of educational evaluation concepts by introducing two evaluation models: the Countenance Model and the Responsive Model. Below is a discussion of these two models:

6.1. The Countenance Model

The Countenance Model, as presented by Stake in 1967, critiques traditional evaluation methods for their failure to provide a comprehensive picture of the subject being evaluated. In response, Stake proposed this model, which identifies two fundamental processes in any evaluation: description and judgment (appraisal). These processes are applied to three stages of an educational program's life: antecedents, transactions, and outcomes.

The stages of an educational program's life are elucidated as follows:

1. Antecedents: These encompass all pre-existing conditions before the program, which may have an impact on it. Therefore, gathering information about these conditions is essential to obtain a thorough background on the program.

2. Transactions: This refers to the activities and events occurring within the program during its implementation.

3. Outcomes: These are the effects that the program has on its participants, resources, and the local community, whether immediate or delayed, direct or indirect.

As for the process of description, it requires gathering information about the three stages: antecedents, transactions, and outcomes to make the description comprehensive. Stake identifies two types of data for each stage: expected (intended) and observed (actual). Expected data are what was planned for in each stage, while observed data are what was actually achieved. After obtaining descriptive information, it is dealt with in one of two ways:

1. Assessing the degree of alignment between the observed and expected for each of the antecedents, transactions, and outcomes.
2. Assessing the extent to which outcomes depend on transactions and antecedents, and the dependence of transactions on antecedents.

Regarding the judgment or appraisal process of the program's life stages, it is conducted for all stages based on relative or absolute criteria. Relative criteria are derived from similar programs, while absolute criteria are defined based on what should be present in the program. In both cases, there must be agreement on the criteria used in making judgments between the evaluator and stakeholders to avoid negative perceptions or attempts to conceal information. Additionally, evaluators can obtain a general assessment of the program's merit by integrating judgments based on absolute criteria with those based on relative criteria.

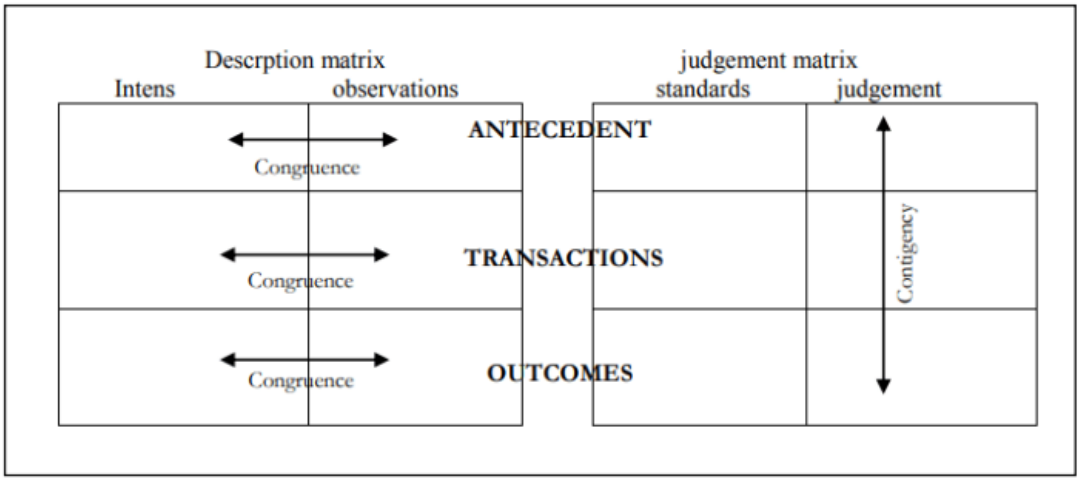
The evaluation process is not complete without providing a justification for assessing the program. Even if the justification is not explicit, it should clarify the philosophical background, fundamental purposes, and the need for the program. Therefore, the steps taken by the evaluator in this model are as follows:

1. Providing a justification for the program.
2. Gathering descriptive information by monitoring the observed and expected aspects of the antecedents, transactions, and outcomes.
3. Analyzing the descriptive information.
4. Identifying absolute and relative criteria.
5. Estimating the program's suitability for those criteria.
6. Issuing a judgment on the program by the evaluator alone or in collaboration with program officials and other stakeholders.

Stake poses some questions that need to be answered before conducting the evaluation because they help in charting the course of the evaluation process, such as:

- Will the evaluation be descriptive, merely describing the program, or evaluative, providing judgments about the program, or both?
- Will the evaluation focus on antecedents, transactions, outcomes, or their functional relationships?
- Will the evaluation emphasize the program's coherence?
- Will the evaluation focus on a specific program, or will it compare it with other programs?
- Will the evaluation aim for development or choosing among alternatives?

Fig 3. Model Evaluation Countenance Stake



Source. Tjabolo & Otaya (2019).

6.2. The Responsive Model

In the early 1970s, Stake began advocating for an evaluation model that addresses the inherent shortcomings in traditional evaluation models, which fail to respond to the issues and concerns of stakeholders (beneficiaries of the evaluation). Stake viewed stakeholders as the ultimate arbiters of evaluation validity.

Based on this perspective, Stake proposed the Responsive Evaluation Model, which relies on inquiries posed by stakeholders regarding the issues and concerns that matter to them. The methodology for gathering information in this model is based on the Comprehensive Evaluation Body Model.

The evaluator in this model ensures to initiate a dialogue with stakeholders to determine the purpose of the evaluation and the type of information needed to address the program's ambiguities. This ongoing communication aims to address stakeholders' inquiries and resolve any challenges they face throughout the evaluation period. Stakeholder engagement continues throughout the evaluation process to accommodate any emergent concerns or demands that were not present at the outset.

According to this model, evaluation is flexible and adaptive, allowing for the incorporation of any new developments. Planning for evaluation begins with a general outline, and customization occurs based on workflow. Evaluation is not predetermined with precise details from the start but evolves during the evaluation process.

In the quest for gathering information, the evaluator ensures to collect data under natural conditions, far from artificiality, through direct observation, engaging in conversations, listening attentively, and relying on multiple sources for data collection. The evaluator also endeavors not to divert stakeholders' attention away from their weaknesses, which they may not wish to disclose when expressing their concerns.

When rendering judgments on the program's quality, the evaluator does not impose specific values; rather, final judgments are derived from the values held by the stakeholders involved in the evaluation process.

7- Hammond Model (Al-Muhaimid, Al-Siriy, & Al-Ajlan, 2014):

The Hammond model focuses on assessing the effectiveness of educational programs in achieving their set objectives. While aligning with Tyler's model for behavior evaluation, Hammond's model goes beyond merely assessing goal achievement to attempting to understand the reasons behind the program's success or failure in meeting its objectives.

The Hammond model entails several sequential steps and procedures to be followed to accomplish the evaluation process and determine the program's achievement of its objectives. These steps include:

1- Identifying the aspects to be evaluated.

2- Identifying the variables related to these aspects.

3- Formulating objectives in operational terms.

4- Evaluating behavior as defined by the operational objectives.

5- Analyzing the evaluation results to determine the extent of achieving each operational objective. This step is one of the additions made by Hammond.

Hammond's interest wasn't solely in defining goals in general and behavior-specific terms; rather, she also focused on elucidating why some objectives succeeded while others did not. Therefore, she designed her evaluation model in the form of a three-dimensional cube, consisting of the educational dimension, the institutional dimension, and the behavioral dimension, as follows:

Firstly, the educational dimension: This dimension concerns the characteristics of the educational program to be evaluated and encompasses five elements:

1-Organization: This encompasses available time, the task schedule, the sequence of topics, the horizontal and vertical organization of the school, and cooperative learning.

2- Content or topics to be covered in the evaluation.

3- Educational activities and interaction patterns in the evaluation among students and teachers, among students themselves, and learning and teaching theories.

4- Facilities and Equipment: This includes transportation, equipment, and essential materials.

5- Cost, encompassing necessary funding, maintenance, and staffing.

Secondly, the institutional dimension relates to the characteristics of individuals and groups associated with the program, including the following elements:

1- Student (age, academic achievement level, family and social background, health, interests, abilities).

2- Teacher (age, educational background, social and family background, personality traits).

3- Administrator (age, educational background, social and family background, personality traits).

4- Educational specialist (age, educational background, social and family background, personality traits).

5- Family (size, income, educational level, socioeconomic status).

6- Local community (geographical location, historical background, economic and social characteristics).

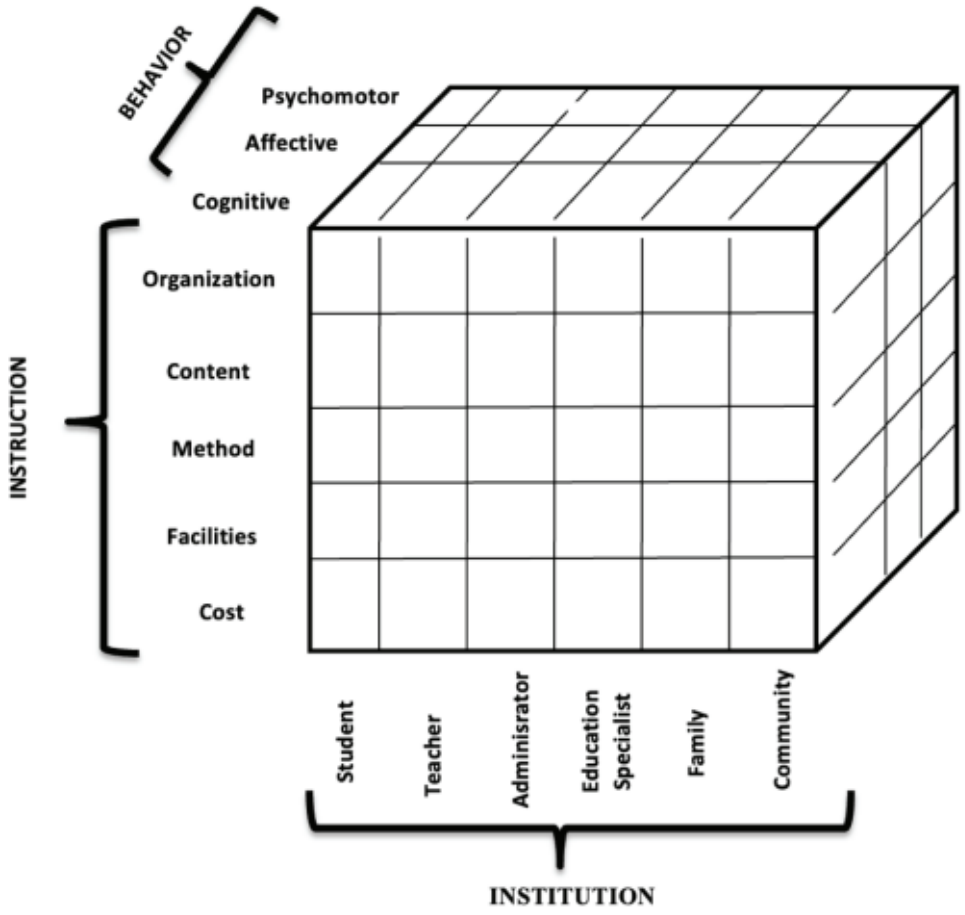
Thirdly, the behavioral dimension includes:

1- Cognitive domain.

2- Affective domain.

3- Psychomotor domain.

Fig 4. Hammond Cube illustrating the interaction of the three evaluation dimensions.



Source: Karagol & Adiguzel (2023).

Thus, Hammond's cube model consists of 90 cells, which is the result of multiplying the dimensions $5 \times 6 \times 3 = 90$. If one were to use this model for evaluation and pose four questions for each cell, it would mean that the total number of questions included in the evaluation would be 360. This is the culmination of the interaction of the three dimensions of Hammond's model with each other (Abu Anab, 2000).

Here's an example illustrating how Hammond's three dimensions interact within his famous cube. If an evaluator addresses the cell resulting from the interaction of content from the educational dimension with teachers from the institutional dimension, along with affective objectives from the behavioral dimension, the questions that can be raised here include:

- How do teachers utilize content to achieve affective objectives?
- What are the feelings of teachers towards the content and affective program objectives?
- Is the program's content sufficient for teachers to achieve their affective objectives?
- Are teachers prepared to use the subjects to accomplish program objectives?

Another example: questions can be derived from the cell resulting from the intersection of the teaching activity element from the evaluation characteristics variable with the student element from the characteristics of individuals associated with the program variable, along with the motor objectives element from the program objectives variable, such as:

1. What is the students' opinion on the program's activities and its motor objectives?
2. To what extent do students participate in teaching activities?

3. Are the current teaching activities suitable to help students achieve motor objectives?

The evaluator is not required to pose questions from every cell, as the cells of interest in a specific evaluation are limited (Abu Anab, 2000).

Educational Assessment in the Algerian Educational System

1- Competency-Based Approach:

In December 2003, the educational curricula in Algeria shifted towards a competency-based approach after the content-based and objective-based approaches. This approach places the learner at the center of the educational process to prepare them to face the challenges encountered globally across all fields. It aims to cultivate citizens who are proud of their identity, preserving their heritage and national culture. Subsequent reforms introduced the second generation of curricula, the Competency-Based Approach, in 2016, which emphasized the social dimension and the student's interaction with their environment. This shift aimed to move from a cognitive constructivist approach to a socio-constructivist one.

The new second-generation curricula focus on competency-based approaches inspired by cognitive, constructivist, and socio-constructivist theories. The foundational approach of these curricula is based on cognitive and socio-constructivist theories, aiming to address the shortcomings of the objective-based approach. Cognitive theory perceives learning as internal cognitive pathways that enable learners to interact with their environment, while socio-constructivism provides strategies for learners to construct knowledge in interactive and meaningful contexts, fostering their contributions to their peer group. Constructivism, on the other hand, emphasizes the importance of constructing knowledge (National Curriculum Committee, 2016a, p. 6).

2- Assessment in the Second-Generation Curricula:

Assessment, as defined by the curricula, is a process of issuing judgments on learning outcomes based on analyzed and interpreted data to make pedagogical and administrative decisions. This assessment is carried out

through its three types: diagnostic, formative, and summative (National Curriculum Committee, 2016b, p. 18).

Assessment has become an integral part of learning, especially formative assessment, which supports learning endeavors among students. Educational assessment serves two main functions: a certifying and evidential function that allows us to assess the level of mastery of targeted competencies and a supportive function aimed at correcting errors and supporting students in their learning. This is achieved through formative assessment by teachers, self-assessment encouraged by students themselves, and peer assessment, which has been highlighted.

Assessment guides the learning process through strategies and efforts to acquire knowledge and solve problematic situations. It also shapes the learning outcomes through knowledge, performance, and emotional behavior. The task of assessment is not only to ensure the acquisition of information but also to make it vital and applicable because success is distinguished by the quality of understanding, competencies acquired, and the quality of knowledge gained, rather than the quantity stored in memory. Therefore, the involvement of students in assessing their work and analyzing it is of utmost importance.

3- Pedagogical Intervention (National Curriculum Committee, 2016b, p. 18):

Pedagogical intervention is the path that enables the learner to overcome the difficulties encountered in their learning process, which appear at several levels within the learning process framework:

1. After the simple learning situation, where the learner's weaknesses or lack of control over knowledge become apparent.
2. After the integration learning situation, where the learner shows weakness in utilizing resources.
3. After solving the initial problematic situation, where the learner demonstrates a deficiency in resource utilization.
4. At the end of the first semester and the end of the second semester, following the results of the interim semester evaluation.

4- Characteristics of Evaluation:

Educational assessment in the second-generation curricula has distinguished itself from the first-generation curricula by features summarized as":

- Evaluating knowledge, efforts, growth of values, and behaviors rather than focusing solely on assessing subject-related competencies and knowledge.
- Formative evaluation is an integral part of learning.
- Utilizing self-assessment and peer assessment.
- Giving importance to pedagogical intervention.
- Reasonable pace of evaluation sessions instead of exaggerating the evaluation sessions.

Thus, Assessment in the Second-Generation Curriculum is characterized by the following:

- Appropriate assessment linked to the second-generation curricula.
- Harmonious assessment with logical and dynamic connections between its components.
- Effective assessment that achieves objectives, facilitates learning, encourages learners to progress in their learning, and truthfully reflects their achievements.

5. Criteria for Assessing Competence:

Competence is defined as the ability to utilize an organized set of knowledge, skills, and attitudes enabling the execution of various tasks. It involves the capacity for action based on the effective mobilization and utilization of resources (acquired knowledge, good conduct, values, intellectual capabilities, personal attitudes, etc.) to address significant problem situations (National Curriculum Committee, 2016a, p. 6).

The criterion or assessment standard refers to the quality that should characterize the student's output, such as precision, clarity, coherence, and authenticity, or the perspective adopted for evaluating any product (National Curriculum Committee, 2016a, p. 24).

The minimum threshold standard is an integral part of competence and serves as the established condition for judging the student's competence. However, the quality criterion is not a prerequisite for accrediting the learner's competence. Authentic solutions and writing styles, for example, are granted additional value in assessing the product, but they do not reject production lacking such qualities.

The student's competence is assessed when they meet all the minimum threshold criteria. We assess the achievement of the minimum standard when a student demonstrates it in two out of three independent examinations that govern it, known as the 2/3 rule.

Minimum Criteria for Assessment

There are numerous minimum criteria widely used and endorsed in the curriculum (National Curriculum Committee, 2016a, p. 25):

1. Criterion of Relevance and Appropriateness: Refers to the conformity of the required product, meaning it does not deviate from the topic and respects the instructions.
2. Criterion of Proper Use of Tools and Resource Deployment: Evaluates the correct use of tools, resource deployment, and application of subject matter rules and concepts.
3. Criterion of Internal Consistency of the Product: Requires a complete, coherent, and reasonable product.

As for the quality or excellence criteria, they pertain to the quality and creativity of the product. However, the value assigned to quality criteria must be limited in consideration of the minimum criteria, according to the 3/4 rule. Quality criteria should not exceed 1/4 of the total score.

6. Assessment Tools:

Educational assessment in the second-generation curriculum employs several tools (National Curriculum Committee, 2016a, p. 24):

1. Assessment Test:

A situation or multiple situations designed to assess student competencies, demonstrating their proficiency against predefined assessment criteria.

2. Assessment Grids:

Observation grids specific to the student and the class, which can be either comprehensive with a single criterion or analytical with multiple criteria, such as rubric-based grids, grading scale grids, and descriptive grids.

- Comprehensive Assessment Grids with a Single Criterion
- Analytical Assessment Grids, which are multi-criteria and include Rubric-Based Grids, Grading Scale Grids, and Descriptive Grids.

3. Monitoring Notebook:

An adaptation of the class notebook containing activities, monitoring cards, remedial suggestions, and summative assessment.

4. Monitoring Card:

A card filled out by the teacher or the students themselves to identify strengths and weaknesses.

5. Assessment Report and School Grading:

Allows parents to evaluate their children's efforts during a semester or academic year.

Achievement Tests

Educational assessment employs measurement tools, which form the basis for assessing academic achievement, whether in specific subjects or across all subjects, and whether over a specified time period or for the entire academic year, in accordance with the educational system in place.

Among the most commonly used and widespread methods for measuring student achievement are achievement tests. These tests determine the extent of a student's progress and motivate them to seek further improvement. They also assist teachers in understanding the extent to which students respond to the learning process, allowing for progression from one level to another. Therefore, achievement tests have garnered significant attention from experts and researchers in the field of education. Consequently, they have developed a set of specifications and conditions that must be considered by the developers of these tests.

Concept of Achievement Tests:

1-1 Definition of Achievement: (Guenoua, 2014)

Before delving into the definition of achievement tests, let's attempt to understand the concept of achievement, which has been defined in various ways. Although there is not significant variation among these definitions, there are some differences worth noting. Haus and Haus define academic achievement as successful or outstanding performance in specific fields of study or endeavors, usually resulting from skill and diligent work and accompanied by attention. This is often condensed into marks, points, grades, or descriptive remarks. (Moulay, 2004, p. 325).

Shaplen defines it as a specific level of performance or proficiency in academic work evaluated by teachers or through standardized tests, or both. (Bro, 2010, p. 207).

Alam defines it as the level of acquisition achieved by an individual or the level of success attained in a specific subject, educational, or teaching field. (Alam, 2002, p. 305).

Mohamed Mustafa Zidan defines it as students' comprehension of lessons and their proficiency in academic subjects, as evidenced by the grades they achieve in exams. (Bro, 2010, p. 208).

Ahmed Zaki Saleh defines it as encompassing everything learned and taught, and since the school's function is to systematically influence students' behavior to bring about specific changes, the content of these changes becomes the subject of achievement. This is because educational goals aim for broad changes in the students' mentality, personality, social relationships, and vocational inclinations, so achievement encompasses all these aspects. (Qureshi, 2002, p. 65).

Based on the provided definitions, we can infer several key elements addressed by these definitions, which are outlined as follows:

- These definitions are limited to what the learner achieves within the intended formal framework according to a specific program aimed at acquiring a set of knowledge and skills. Academic achievement does not encompass unintentional, informal learning.

- Some definitions emphasize success and complete performance, but it's worth noting that academic achievement, in general, can vary from good to weak or poor and may not always reach the level of success and completeness.

- The definitions also highlight that academic achievement is always associated with the curriculum presented or prescribed, which may consist of either a subject or a group of subjects within a specified time period.

- The definitions emphasize that measuring academic achievement requires mechanisms to assess its level, such as achievement tests, which may be prepared by the teacher or standardized tests.

- Academic achievement is what has actually been attained and accomplished, representing the learner's final outcome in a specific academic field, rather than merely latent ability or preparedness.

1-2 Definition of Achievement Tests:

An achievement test is a method or means through which a series of tasks is presented to the examinee, who is expected to respond to them in a manner that indicates the extent of their possession of that attribute. (Abdelhadi, 2001, p. 277)

An achievement test is defined as "the tool used to measure knowledge, understanding, and skills in a specific academic or instructional subject, or a group of subjects." (Al-Faqih, 2005, p. 464)

Additionally, an achievement test is defined as "an organized procedure to determine the extent of student learning." (Malham, 2001, p. 433).

Airasian (1997) defines the achievement test as "an organized method to determine the level of student achievement of information and skills they have learned." (Qattami, 2009, p. 354)

Furthermore, Saadah defines the achievement test as "an organized procedure in which observations of students' behavior are made to ensure the achievement of set objectives through the presentation of a series of items or questions that require responses, with these responses described using numerical scales." (Abu Jaoud, 2005, p. 411)

Brann defines the achievement test as one that measures learning that occurs as a result of experience in a controlled learning situation, such as a classroom or training program. The frame of reference of interest is the present or the past, meaning the current state of what students have learned. (Alam, 2006, p. 123)

From the aforementioned definitions, we can ascertain that an achievement test is an organized method to determine the level of students' achievement of previously learned information and skills in a specific academic subject through their responses to a set of items representing the content of the subject matter.

2- Specifications of Achievement Tests:

Ensuring the quality of educational assessment requires examining the good specifications that must be present in assessment tools and achievement

tests, which are among the most important assessment methods. These specifications must be clear to both teachers and students alike. They include:

2-1 Validity:

Refers to the extent to which the test measures what it is intended to measure, making it suitable for use in light of the intended objectives. If a test is designed to measure students' achievement in mathematics, for example, it should focus on assessing mathematical abilities rather than their attitudes towards mathematics. Content validity is emphasized, ensuring that the test content reflects what the teacher has taught in the subject matter.

2-2 Reliability:

Refers to the consistency of the relative position of a student, meaning that their scores remain stable if the test is repeated on the same student or equivalent forms are administered to the same group of learners. Reliability can be calculated using various methods recognized by psychological and educational measurement specialists.

2-3 Objectivity:

Ensures that learners' results are not influenced by the subjective biases of the scorer, meaning that personal factors do not affect the determination of learners' grades, even in cases of multiple scorers. Objectivity also ensures consistency in grading, even when multiple scorers are involved.

2-4 Comprehensiveness:

Ensures that the test adequately covers the educational objectives or competencies to be measured, as well as the content of the subject matter. This can be achieved by analyzing the content and linking it to educational objectives or competencies while determining the relative weights based on importance or the required time. This process, known as test blueprinting, helps in designing questions and grading scales accordingly.

2-5 Discrimination:

A good test is one that can discriminate between different levels of students' achievement, meaning that their scores vary according to their levels of achievement. For example, a student with low achievement would receive a low score on the test, while a student with average achievement would receive

a moderate score, and a student with high achievement would receive a high score. Statistically, there should be a noticeable difference between the averages of the lowest and highest groups.

2-6 Applicability:

Tests should be designed to be easily applicable, especially for assessments involving large numbers of examinees. This consideration prevents unnecessary expenditure of effort, time, and money.

2-7 Ease of Grading:

Some add this criterion because, sometimes, although a test is well-constructed according to the aforementioned specifications, grading it may be complex and difficult.

2-8 Clarity of Instructions:

A crucial aspect of a good test is that the instructions are clear and easily understandable to students, without ambiguity or the need for further explanation, and without resorting to traps to confuse students.

2-9 Sufficient Time for Learners:

The allotted time for students to answer the test must be carefully considered. Providing sufficient time ensures greater validity, but it should not exceed its limit significantly.

3.Types of Questions in Achievement Tests

Achievement tests can be either oral or written, but the most widely used are written tests, employing the paper-and-pencil strategy. In this format, learners are assessed using a provided paper containing a set of questions, which they answer in writing. Typically, this occurs under controlled conditions, particularly with respect to timing.

3-1 Essay Questions:

These are questions that require students to provide free-form responses to a given topic or problem. They are used to assess educational objectives that emphasize the importance of expression, such as comparing two things, forming and defending opinions, explaining meanings, critiquing ideas, analyzing, suggesting problem solutions, understanding, critiquing, and evaluating. Essay

questions vary in the degree of freedom allowed in the response and are classified into limited-response essay questions and open-response essay questions. Essay questions are advantageous in that they encourage good learning habits, such as the ability to form and defend opinions, compare and justify, summarize, analyze, reorganize, explain, problem-solve, understand, critique, and evaluate. However, they also have drawbacks, such as lower reliability due to the inability to represent the entire curriculum content, the potential influence of luck on a student's mastery of the test topic, and the difficulty and subjectivity of grading.

3-2 Objective Questions:

These are questions that require short, specific responses, with each question having a single correct answer. They are called objective because grading is done objectively, meaning they do not rely on the grader's subjectivity in evaluating the answer. Instead, they rely on a standardized answer key used by all graders for the same question.

Types of Objective Questions:

- True/False Questions: These consist of statements or factual sentences, requiring the student to judge their accuracy.
- Multiple Choice Questions: These comprise a stem followed by a set of options or proposed solutions, with the student tasked to identify the correct choice. This type may have either one correct option or a set of choices, depending on the question.
- Completion Questions: These consist of incomplete sentences or questions, prompting the student to fill in the missing parts with a word, phrase, or symbol that fits the context.
- Ordering Questions: In this type, the student is given a set of scattered phrases and asked to arrange them according to a specific criterion.
- Matching Questions: These involve two parallel lists of phrases, requiring the student to match each item from the first list with a corresponding item from the second list.
- Classification Questions: The student is given a set of phrases and asked to classify them into lists based on a common property.

Objective tests are characterized by having a large number of questions covering the content of the subject or curriculum, which enhances the content validity as students must master the entire material, reducing the element of chance. However, they require significant effort and time to prepare, in addition to the necessity of having the ability and skills on the part of the teacher to construct objective achievement tests, along with training on how to do so.

3.3. Composite Questions:

Composite questions focus on the unity of the topic and consist of several paragraphs with varying levels of difficulty and skills being assessed. When constructing composite questions, the same considerations as other types of questions should be taken into account. These questions can be presented in one of the following forms:

• Problem-Based:

The question is formulated in a way that enables the student to utilize the information provided in the question and their existing knowledge of the topic to solve a problem. The student employs this information and prior knowledge to address the problem presented in the question.

Benefits of composite questions:

- They assist students in transitioning from one level to another within the same topic as students organize their responses.
- Students focus their responses on specific objectives within the given topic.
- They accommodate the varying levels and abilities of students, allowing them to respond to the question regardless of its complexity. (Al Rawashdeh et al., 2000, 7).

4- Constructing Tests (Guenoua, 2015)

4-1 Educational Objectives and Their Classification:

The educational process, like any other human activity, revolves around delineating a set of objectives and striving to achieve them. These objectives represent the ultimate outcomes of this process. Bloom et al. perceive the objective as an attempt by the teacher or curriculum specialist to explore the

variables that occur in the learner after experiencing an educational experience (Malham, 2001, p. 112). Similarly, Bashir Maamria describes the objective as "the expected change in the learner's cognitive, affective, and psychomotor behavior after receiving a specific educational content" (Maamria, 2007, p. 138).

Therefore, educational objectives target aspects of the learner. In this regard, Benjamin Bloom and his colleagues worked on classifying educational objectives as follows:

4-1-1 Cognitive Domain:

Bloom and others classify educational objectives in this domain into six levels arranged hierarchically from the lowest level at the base of the hierarchy to the highest level at the top. These levels are as follows:

1. Knowledge:

This involves knowing things, facts, terms, theories, and concepts. Anwar Aql believes that questions at this level begin with one of the following verbs: define, identify, label, list, name, or select (Al-Atoum, 2005, p. 44).

2. Comprehension:

At this level, the learner can perceive the studied material. This is evident through their ability to translate texts and figures, interpret ideas, and complete information. Verbs associated with this level include describe, explain, construct, illustrate, summarize, discuss, etc.

3.Application:

This level measures the learner's ability to use what they have learned in new situations. This is reflected in their ability to apply rules, theories, concepts, and methods correctly in new scenarios. Questions at this level are associated with the following verbs: apply, implement, organize, solve, provide examples, draw, etc. (Maamaria, 2007, p. 142).

4.Analysis:

Norman Grounlund states that analysis refers to the learner's ability to break down the learned material into its components to help understand its structural organization. Questions at this level often begin with verbs such as analyze, describe, distinguish, identify, differentiate, compare, calculate, deduce, etc. (Al-Atoum, 2005, p. 44).

5.Synthesis:

This refers to the learner's ability to put parts together to create something new, such as preparing a research project, presentation, or delivering a specific lecture (Ma'amaria, 2007, p. 141). Grounlund believes that the learning outcomes at this level emphasize creative behavior. Actions associated with this level include: combine, organize, modify, create a plan, generalize, narrate, etc. (Al-Atoum, 2005, p. 44).

6.Assessment:

This is the mental process through which the learner makes judgments based on specific criteria. At this level, the learner forms a judgment about the value of a subject or thing, based on a certain set of criteria. Verbs associated with this level include criticize, evaluate, give your opinion, compare, balance, prove, etc. (Al-Noor, 2007, p. 61).

4-1-2- Affective Domain:

According to Al-Atoum (2005), affective objectives aim at developing positive attitudes, values, habits, interests, good inclinations, and positive emotions and regulating and directing them effectively. The affective domain encompasses emotions and sentiments expressed through attitudes, inclinations, interests, acceptance, preferences, and other behaviors that reflect individuals' feelings and emotions. Therefore, affective objectives aim to develop and regulate them positively and effectively. Additionally, these objectives are characterized by their generality and the difficulty of formulating them specifically. They are challenging to measure and control the extent of their achievement. Hence, teachers often overlook them and focus more on cognitive objectives.

The classification of affective educational objectives was introduced in 1964 by David Krathwohl and others, following the same hierarchical approach as Bloom's taxonomy, ranging from simple to complex. The affective domain is categorized hierarchically into five levels:

Receiving:

Receiving implies the learner's sensitivity to the existence of phenomena or stimuli, being receptive to them, distinguishing them, and preferring them over

other stimuli. It includes three sub-levels: awareness, willingness to receive, and selective attention or preference.

Responding:

It refers to the learner's responses and reactions to stimuli and topics received. It includes three sub-levels: compliance, willingness, and satisfaction. (Ghanem, 2002, p. 37).

Valuing:

Giving value to a subject or behavior: At this level, the learner's activities and attitudes are linked. The learner's intellectual beliefs, preferences, commitments, and adherence determine their responses. According to Fuad Abu Hatab (1979), this level includes three sub-levels: acceptance, preference, and commitment. (Maamria, 2007, p114).

Organizing Values into a Structure or System:

At this level, the learner integrates specific values into a unified value structure or system. Benjamin Bloom (1983) suggests that within this level, there are two sub-levels: perception or conception of a specific value and organizing the value structure (Murad & Amin, 2005, p. 117).

Possessing a Compound of Values:

At this level, the learner has established a stable value system that governs their behavior and gives them a distinct philosophy and lifestyle. Others can predict their behavior in certain situations. Two sub-levels fall under this level: general orientation and predisposition or disposition. (Al-Atoom, 2005, p. 45).

4-1-3-Psychomotor Domain:

This domain includes educational objectives that emphasize skills requiring coordination between the members of the motor system and the nervous system. It encompasses motor skills such as running, jumping, and rotating, as well as manual skills like writing, drawing, and playing instruments. Elizabeth Simpson's classification in 1972, following Bloom's approach, is one of the most renowned in this domain. She categorized these objectives into a hierarchical organization from simple to complex, comprising seven levels:

Perception:

At this level, sensory organs are used to obtain roles leading to motor activity. It ranges from mere sensory stimulation to selecting roles related to it and linking roles to performance to perceiving the tools and objects used in motor tasks.

Set:

This refers to the learner's readiness to perform a specific movement, encompassing both physical and mental inclinations as well as emotional inclinations. Each of these three inclinations influences the others. For example, if a learner has a mental inclination to perform a certain movement, the mind may send a signal to the limb ordering it not to perform the movement due to the influence of emotional inclination on mental inclination and subsequently on physical inclination and the level of readiness.

Guided Response:

This level focuses on the initial stages of performing complex movements. These stages start with imitation, where the learner imitates a specific movement performed by the teacher, and then move to trial and error, where the learner performs multiple movements to determine the correct or optimal one or practices the movement to improve performance later.

Mechanism:

At this level, learned responses have become habitual, and the movements can be performed with some confidence and proficiency.

Complex or Overt Response:

In this level, complex movements are performed with a high degree of precision and skill, reaching a certain level of proficiency.

Adaptation or Modification:

At this level, the learner can adjust movement patterns to fit specific requirements. Here, they have become familiar with the movement, mastered its details, and practiced it with high accuracy and speed.

Origination:

This level focuses on creating new patterns to fit a particular situation or specific problem. Salah Ahmed Murad and others add that the learning outcomes at this level emphasize creativity based on highly developed skills,

meaning the learner innovates new, advanced movements to suit the new situation they are dealing with.

4-2 Steps in Constructing an Achievement Test:

Building a good achievement test requires adherence to specific steps that embody the following:

4-2-1 Defining the Purpose and Objectives of the Test:

In this step, the intended goal of conducting the test is specified (Ibrahim, 2008, p. 56). For example, if the purpose is diagnostic, to identify areas of weakness, the test is designed accordingly. Additionally, it is essential in this step to identify the topics and learning outcomes to be measured among learners to ensure test coverage and inclusivity.

4-2-2 Preparation of the Specifications Table:

Anastasi (1988) perceives the specifications table as comprising two dimensions. The first dimension denotes levels of cognitive objectives such as understanding, application, analysis, etc., while the second dimension represents units or content of the study material. Through this table, the instructor can dissect the study material into subheadings, determine the cognitive objective levels for each part of the content, and ascertain its relative importance. Consequently, the instructor can select test items representative of the study material content and objectives from various cognitive levels (Al-Zaghloul, 2009, p. 325).

Table 1 : Model of the Specifications Table in Light of Cognitive Levels. Content

Educational Objectives / content	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	relative weight percentage
Subject1							
Subject2							
Subject3							
Subject4							
relative weight percentage							

Table 2 : Model of the Specifications Table in Light of Cognitive, Affective, and Psychomotor Levels.

Educational Objectives / content	Cognitive processes	Affective processes	Psychomotor processes	relative weight percentage
Subject1				
Subject2				
Subject3				
Subject4				
relative weight percentage				

For the construction of the specifications table to be accurate, the teacher should follow the following steps as outlined by Al-Nour (2007):

- The teacher identifies the study topics to be assessed.
- Determination of the necessary number of sessions required to teach each topic.
- Determination of the relative weight of the study topics through the following equation:

$$\frac{\text{Number of classes required to teach the topic} \times 100}{\text{Number of classes needed to teach the entire subject}} = \text{Number of classes for the importance of the topic}$$

- Identifying the behavioral objectives to measure the extent of their achievement by the student at different levels and determining their relative weights using the following equation:

$$\text{Relative weight of goals at a given level} = \frac{\text{Number of goals at that level} \times 100}{\text{Total objectives of the full course}}$$

- Determining the total number of test questions based on the allowed time for response and other variables related to student characteristics.
- After determining the total number of questions, the number of questions for each topic at each level of objectives is determined using the following equation:

Number of topic questions = total number of questions x relative weight of topic objectives

- Determining the scores of questions for each topic at each level of objectives utilizing the following equation:

Subject Question Score = Final Test Score x Relative Weight of Topical Relevance x Relative Weight of Subject Objectives

- - Adhering to all the aforementioned steps in preparing the specifications table ensures comprehensiveness, which is a key characteristic of a good test. This is achieved through the fair distribution of questions across topics according to their respective importance, thereby preventing focus on a small portion of the

studied subjects. Consequently, the questions cover all aspects of the curriculum.

4-2-3 Test Formulation

After determining the distribution of questions across topics and objectives, we proceed to identify the appropriate types of items for each topic, whether they are essay-based, objective, or a combination of both. We also specify the types of objective questions, such as ordering, matching, multiple-choice, short-answer, or brief essays.

Next, we formulate each question within the specified topic and objective, ensuring simplicity, clarity, ease of expression, and linguistic accuracy while avoiding complexity or misleading elements that could hinder understanding of the required task.

Subsequently, we draft the test instructions, clearly stating what is required to answer the questions and how to do so while coordinating between different sections and standardizing instructions whenever possible. We also consider the allotted time for the test, ensuring it is sufficient for a good examination, and we articulate this in the instructions.

Upon formulating the questions and instructions, we produce the preliminary version of the test, paying attention to formatting aspects such as font type, size, and paper size, along with providing model answers for its items.

4-2-4 Test Trial and Finalization

This step occurs after completing the test design, where it is applied to a random sample resembling the group to which the test will be administered. Appropriate statistical analyses are then conducted to assess the test's reliability, validity, objectivity, and discriminative power. Based on these analyses, necessary adjustments are made accordingly. (Al-Zughoul, 2009, p. 325).

Building a good achievement test involves a sequential process, starting with defining its objectives and purpose, culminating in its initial form, followed by trial administration for statistical analysis. Any necessary modifications are then made if deficiencies are identified. Diligently attending to each step of the preparation process significantly contributes to developing an achievement test founded on sound scientific principles, meeting the required criteria for

construction, and ultimately providing a valid assessment tool capable of issuing accurate judgments and making informed decisions.

5. Analysis of Test Items:

Analyzing test items is a procedure used to assess the quality of test items in terms of their difficulty level, discriminative power, and the effectiveness of each alternative in multiple-choice questions (Khattab, 2000, p. 95).

To calculate these indicators, it is necessary to analyze students' responses to test items or questions. The goal of this analysis is to evaluate the items themselves, regardless of assessing the students, to provide teachers with insights into the quality of the items.

Analyzing test items by the test developer transforms them into proficient item writers. Through practice and training, many defects affecting the items, whether in formulation or technical errors, can be eliminated. This process enables the creation of items with appropriate difficulty levels and discrimination indices, which can be retained and utilized in future assessments.

Furthermore, item analysis not only reveals the difficulty and discrimination levels but also assesses the effectiveness of alternatives (distractors). Alternatives that are chosen more frequently by the higher-performing group undergo review, as do those that none of the students selected or were selected by a proportion below the required threshold (Aziza, 2016, p. 66).

5.1. Difficulty Index:

The difficulty index refers to the proportion of students who answered a question correctly to the total number of students who took the test, ranging in value from zero to one (Khattab, 2000, p. 95).

Some may refer to this index as the ease coefficient, assuming that its elevation indicates the ease of the question rather than its difficulty. However, it is correctly termed the difficulty index.

Calculation of the Difficulty Index for Objective Questions:

$$\text{Difficulty Index} = \frac{\text{number of correct answers}}{\text{Number of respondents}} \times 100$$

Example: If we have 50 students, and 24 of them answered a certain question correctly, then the difficulty index for this question is.

$$n\ y = (24 \div 50) \times 100 = 0.48 = 48\%$$

As for essay questions, the difficulty index can be calculated using the following equation: (Alam, 2015)

$$\text{Difficulty Index} = \frac{\text{Total scores obtained on the question}}{\text{Number of students} \times \text{question score}}$$

Example: If 20 students answered an essay question in history, with a maximum score of 5 points, and their total combined score on the question (the sum of the scores they received) is 75 points, then the difficulty index is:

$$\text{Difficulty Index} = \frac{75}{20 \times 5} = \frac{75}{100} = 0.75 = 75\ \%$$

The answers to multiple-choice questions may be influenced by guessing and chance, and in a manner that does not indicate an increase in achievement. This influence increases as the number of options in the question decreases.

Therefore, I propose correcting the difficulty index for the effect of guessing using the following equation by Guilford (Mokadem, 1993, p. 139):

$$\text{Corrected Difficulty Index} = \frac{B \times N - 1}{B - 1}$$

Where:

- (N): The obtained ratio of the difficulty index.
- (B): The number of options.

Example: If the difficulty index of a question is 60% and it is a multiple-choice question with 4 options, then the corrected difficulty index equals:

$$\text{Corrected Difficulty Index} = \frac{4 \times 0.60 - 1}{4 - 1} = \frac{1.4}{3} = 0.46 = 46\%$$

As the difficulty index increases, it indicates the question's ease, whereas a decrease suggests its difficulty. If the difficulty index is 100%, it means the question is very easy and all students answered it. Conversely, if the difficulty index is 0%, the question is extremely difficult, and not all students answered it. A difficulty index of 50% is considered good, with values between 40% and 60% being acceptable and desirable. (Maamria, 2007, p. 126)

The quality of the test requires diversifying question levels in terms of difficulty and ease, increasing the number of questions, and covering the subject matter comprehensively. Additionally, minimizing cheating and guessing should be prioritized to prevent students from receiving undeserved scores. This can be achieved by including easy questions (16% of the test) to accommodate weaker students, moderate questions (68%) for average students, and difficult questions (16%) for high-achieving students. Easy questions don't imply that no student might fail, and difficult questions aren't impossible but require substantial mental effort. Starting with easy questions in the test and progressing to difficult ones stimulates student motivation for engagement (Al-Nour, 2007, p. 207).

5. 2. The Discrimination Index:

Mokadem (1993) defines the discrimination index as the extent to which an item can distinguish between two groups. The primary objective of most tests is to differentiate between individuals who were tested. A question that effectively distinguishes between high-achieving and low-achieving students is said to have discrimination ability. There are multiple methods for calculating the discrimination index, two of which are outlined:

The first method involves calculating the discrimination index of the question by computing the correlation coefficient between students' scores on the question and their total scores on the entire test. It can also be calculated by determining the correlation coefficient between the question scores and an external criterion that aligns with the test's purpose. The Pearson or Spearman correlation coefficient can be chosen based on the data.

The second method, which is more commonly used, relies on comparing the differences between the upper and lower groups. Students who answered the test are divided into upper and lower groups after arranging them in descending or ascending order based on an external criterion or the total test score. The upper group comprises 27% or one-third of the students if the number is relatively large, or half (50%) if the number is small.

An example of this would involve the following steps:

1.Sorting the papers in descending order according to the grades, assuming there are 100 papers.

2.Dividing the papers into two groups: the upper group, representing the top 27% of papers with the highest grades, and the lower group, representing the bottom 27% of papers with the lowest grades. Thus, the number of individuals in both the upper and lower groups is 27.

3.Counting the number of students who correctly answered the first question among those who obtained high grades. Let's assume there were 20 students.

4.Counting the number of students who correctly answered the first question among those who obtained low grades. Let's assume there were 12 students.

5. Calculation of the index of discrimination.

$$\text{Discrimination Index} = \frac{\text{Number of correct answers from the upper group} - \text{Number of correct answers from the lower group}}{\text{Number of individuals in one group}}$$

For example:

$$\text{First question discrimination index} = \frac{20-12}{27} = 0.29 = 29\%$$

Calculation of the discrimination index for essay questions:

$$\text{Discrimination Index} = \frac{mg\ x - mg\ y}{\text{Mug m} \times N}$$

Where:

- (mg s): Total grades obtained by the upper group of students.
- (mg r): Total grades obtained by the lower group of students.
- (mg m): Grades allocated to the question.
- (N): Number of individuals in either group.

Example: Let's assume the total grades obtained by the upper group of students for the first question in a certain subject, for example, are 70 points, and the total grades obtained by the lower group of students for the same question are 38 points. The number of individuals in both the upper and lower groups is 8, and the score allocated to this question is 10 points. Calculate the discrimination coefficient.

$$\text{Discrimination Index} = \frac{70-38}{10\times8} = \frac{32}{80} = \frac{4}{10} = 0.4 = 40\%$$

The discrimination coefficient is bounded between -1 and 1. This coefficient is considered good when it equals one and satisfactory if it reaches 0.60 or 0.70. The discrimination coefficient appears in three conditions or degrees: it can be high, low, or negative.

There is no rule for accepting a question or item based on discrimination, but according to Alam (2006):

- Any item (question) with a negative discrimination index is deleted.
- Any item (question) with a discrimination index from zero to 19% is considered weak in discrimination and is also recommended for deletion.
- Any item (question) with a discrimination index between 20% and 39% is considered acceptable in discrimination and is recommended for modification.
- Any item with a discrimination index above 39% is considered strong and good in discrimination.

5. 3. Effectiveness of Alternatives:

In multiple-choice questions, the options represent possible alternatives for answering, with one option being the correct answer. It is assumed that the other options represent plausible answers and should attract learners. There must be selection of any incorrect alternative by at least one student or at a rate of not less than 5% of the learners, with the majority being from the lower-performing group.

Options that do not attract anyone or attract very few of them are considered ineffective alternatives and should be replaced. Options that attract more students from the higher-performing group than from the lower-performing group are eliminated. This indicates an opposite direction to the item's discrimination. The more negative the value of the alternative, the better the distractor (Aziza, 2016).

Some students resort to guessing their answers when they are unsure of the correct answer. Therefore, the test instructions should clarify the possibility of guessing in such cases, or the question may be left unanswered in tests that prohibit guessing, which may result in negative marking. Thus, it is preferable for the instructions to encourage thoughtful consideration when answering all items. If one is unable to answer a specific question, they may resort to guessing based on partial information rather than leaving the answer blank, as intelligent guessing is an important skill practiced in daily life and is encountered in numerous situations (Alam, 2015).

Performance Assessment

1. Definition:

Performance assessments have evolved beyond solely focusing on cognitive aspects to encompassing emotional and kinesthetic dimensions. They evaluate competencies that cannot be measured through traditional oral or written tests, instead emphasizing practical performance and skills. Performance assessments measure individuals' ability to translate acquired knowledge into practical scientific situations and to demonstrate scientific thinking skills, problem-solving abilities, manual skills, and the ability to control the use of scientific devices and tools (Yehia, 2017, p. 25).

Performance assessments are used to gauge students' ability to translate acquired scientific knowledge into practical scientific situations. They also measure scientific thinking skills, problem-solving abilities, manual skills, and the ability to control the use of scientific devices and tools (Taabli & Gouareh, 2013, p.187).

Performance assessments encompass both performance tasks and products. Evaluating performance tasks requires students to demonstrate their skills in practical operations related to educational objectives under specific conditions. Both the teacher and the student observe the performance. On the other hand, evaluating products relies on the student's completion of a task that requires producing a specific outcome, such as writing a research report, completing a project, drawing, presenting an exhibition, or developing a computer program (Alam, 2015, p.139).

2. Types of Performance Assessments:

Performance assessments aim to measure learners' performance and their actions and productions. They focus on skill requirements and are used in

various fields such as physical and natural sciences, vocational training programs, computer literacy, and language learning.

The importance of performance assessments lies in enabling learners to recognize their actual performance skills in real-life situations. There are three types of performance assessments:

2-1 Matching Test:

This type of assessment includes various test situations that present different degrees of realism. Sometimes, the teacher asks about the definition of a machine or a piece of equipment, then points out its functions and benefits. Alternatively, learners may listen to the sound of a malfunctioning machine and, based on the sound, identify the cause of the malfunction and suggest appropriate steps for repair.

2.2 Imitation Performance Test:

This test requires learners to replicate specific steps, expecting them to perform and mimic the same movements. Examples include tests where learners are required to create something or perform specific actions such as prayer rituals.

2.3 Sample Work Test:

This type of assessment epitomizes the highest degrees of realism, requiring the learner to perform realistic tasks representative of the overall performance being measured. The task sample encompasses most elements of the overall performance and is conducted under control and supervision (Hanouda, 2013, p. 113).

3. Performance Assessment Methods (National Assessment Team, p. 37):

Performance assessments can be conducted through various methods, including:

3.1 Presentation:

A structured and organized presentation conducted by the learner or a group of learners on a specific topic within a specified timeframe to demonstrate their mastery of specific skills. The learner(s) would provide an explanation of a topic supported by techniques such as images, diagrams, and electronic slides.

3. 2 Explanatory Presentation:

A verbal or scholarly presentation conducted by the learner or a group of learners to clarify a concept or idea, demonstrating the learner's ability to reiterate the concept in a clear manner and language. The learner would elucidate a concept through a scientific experiment or by relating it to real-life situations.

3-3 Practical Performance:

A set of actions aimed at demonstrating knowledge, skills, and attitudes through the learner's execution of specific tasks practically. Learners are tasked with using a device, designing a computer program, car engine maintenance, hair styling, fashion design, administering injections, preparing a dessert dish, or producing a model, map, prototype, or product.

3-4 Speaking:

The teacher or a group of learners speaks on a specific topic within a short and defined period, often narrating a story, retelling a narrative, or presenting an idea to demonstrate their ability in expression, summarization, and connecting ideas. For instance, learners might speak about a movie they watched, a trip they took, a story they read, or summarize ideas discussed in an educational setting to convey them to another group.

3-5 Exhibition:

Learners showcase their intellectual and scientific output in a designated location and time to demonstrate their ability to employ their skills in a specific field to achieve a particular outcome. This may include displaying models, prototypes, photographs, paintings, artworks, products, fashion items, or handicrafts.

3-6 Simulation or Role-Playing:

The learner engages in dialogue or discussion accompanied by movements and gestures required by the role in a situation resembling a real-life scenario to demonstrate their cognitive and performance skills, as well as their ability to follow instructions, communicate, make suggestions, and make decisions through a task or problem-solving. The scenario can be technically simulated,

where the learner integrates into a calculated simulation scenario and is required to execute the same type of tasks and decisions expected in their future work. While simulation scenarios relying on learner explanation present scenarios with human or non-human elements, technically based simulation scenarios present a scenario on a computer screen where the learner can demonstrate their decision-making abilities. The program provides hundreds of different scenarios and elements.

3-7 Discussion or Debate:

A meeting between two teams of learners for argumentation and discussion about a certain issue, where each team adopts a different viewpoint. Additionally, there is an adjudicator (one of the learners) to demonstrate the learners' ability to persuade, communicate effectively, actively listen, and present arguments and justifications supporting their viewpoint.

4. Characteristics of Performance-Based Assessment (National Assessment Team, p. 38):

- Direct assessment of roles as they are in real life or simulated, where cognitive, performance, and emotional skills are assessed, thereby deriving its credibility and authenticity.
- Comprehensive assessment focusing on both processes and outcomes.
- Enables learners to play a positive and active role in searching for information from multiple sources and processing it.
- Allows learners to engage in self-assessment during the execution of a task, work, or project.
- Involves learners and teachers in setting performance assessment criteria and performance levels based on these criteria.
- Provides learners and teachers with the opportunity to modify assessment procedures and tasks based on feedback from either party, thereby encouraging learners to achieve a high level of quality.
- Provides learners with the opportunity to defend their performance with logical and practical arguments and evidence.

5. Steps for Designing Performance-Based Assessments:

1. Identifying the desired performance objectives.
2. Determining the tasks to be used, whether they are performance-based or product-based.
3. Specifying the conditions under which the tasks will be performed.
4. Preparing instructions for students and writing a list of skills and processes to be evaluated, along with specifying the time frame for completion.
5. Establishing criteria and performance levels.
6. Reviewing the tasks and making revisions if necessary (Alam, 2015, p. 140).

6. Constructing Criteria for Evaluating Performance:

Constructing Performance Evaluation Criteria involves reaching standards and criteria for performance that can be observed or judged by specifying the performance, product, or task to be evaluated. This includes creating a list of performance or product characteristics and identifying the most important ones to reduce the number of criteria for easier observation. These criteria should be expressed in observable student behaviors or product characteristics, clarifying and arranging them according to their expected appearance (Alam, 2015, p. 147).

These criteria or standards are then placed in evaluation grids in the form of scales of assessment, which may be analytical with many criteria or holistic through comprehensive evaluation based on a single criterion. These grids are similar to observation grids, but the difference lies in the fact that observation is direct and not known to the student, whereas performance testing occurs under specific conditions where the student is aware that it's an assessment.

It is essential to note that it is better to involve students in the development of these criteria, in addition to clearly informing them about them.

Example: An Example of Performance Assessment Criteria for a Semester Research Paper/Presentation of the TD of the Educational Assessment Module, Prepared by Dr. Abdellatif Guenoua and his students.

1- **Personal Information**

	Evaluation Card		
Date		**Group:**	
The title			
The Presenters			
Evaluator			

2- Evaluation of the Theoretical Aspect through the Submitted Written Research:

	Theoretical Aspect (Written)				
	Weak	**Average**	**Above average**	**Good**	**Excellent**
The Plan					
The Introduction					
The Content					
The Conclusion					
Language Accuracy					
Citations					
Presentation					
Organization					
Applied Aspect					
Adherence to the Time and Method of Submission					
Resources					

3- Presentation Assessment:

	Presentation				
	Weak	**Average**	**Above average**	**Good**	**Excellent**
Planning					
Explanation					
Organization of Information					
Time Management					
Language Accuracy					
Interaction					
Voice					
Eye Contact					
Use of Visual Aids					
Summary					
Answering Questions					

4- Evaluation of the Applied Aspect (Role-Playing)

	Applied Aspect (Performance)				
	Weak	**Average**	**Above average**	**Good**	**Excellent**
Relevance of the Topic					
Content					
Organization					
Documents and Attachments					
Tools					
Introduction to the Application Site					
Consistency					
Performance					

Observation

1- Definition of Observation:

Observational assessment relies on gathering information about learner behavior and verbally describing it. It is typically a qualitative form of assessment, providing direct evidence of learners' learning. It is a process in which the assessor directs their various senses towards the learner in an active situation to obtain information useful for judging them and evaluating their skills, values, behavior, ethics, and thinking style.

2- Types of Observation:

2-1 Simple Observation: The observer spontaneously observes behaviors as they naturally occur in real-life situations.

2-2 Systematic Observation: This type of observation is planned and tightly controlled, with pre-established plans determining the time, place, and criteria.

3- Characteristics of Observation:

- Provides information about some learning outcomes that cannot be obtained through other methods.
- Offers quantitative, qualitative, and comprehensive information, enhancing confidence in assessment results.
- Possesses high flexibility, allowing adaptation and design to suit different outcomes.
- Provides information about learners' abilities in real-life situations.

4- Benefits of Observation:

- Provides students with a clear picture of their development.
- Offers a broad scope for assessment.

- Allows teachers to create a conducive environment by observing certain trends and skills.
- Helps identify problems as they arise.
- Provides immediate feedback.

5- Steps for Designing Observation:

1- Determine the purpose.
2- Identify learning outcomes.
3- Specify the required practices, tasks, and performance indicators.
4- Arrange practices, tasks, and indicators in a logically sequenced table.
5- Design a recording tool for these practices and indicators.

6- Types of Observation Grids:

6-1 Analytical Grids:

6-1-1 Single-Aspect Observation Grids (Rating Scale):

These grids resemble checklists in their content, as each statement or question listed in the grid aims to observe behavior and is assigned several response options. Typically, there are five options in its structure, such as (always, often, sometimes, rarely, never) or (excellent, very good, good, acceptable, poor). Here, the teacher can accurately assess the desired trait by assigning a judgment to it. (Maamria, 2012)

Example:

Criteria and indicators	**Very easily**	**Easily**	**Barely**
Knows vocabulary			
Uses teaching aids			

6-1-2 Binary Grids (Checklist):

These grids consist of a set of paragraphs that the teacher places to measure a certain trait in the student's behavior. These paragraphs, which focus on some simple behaviors, are arranged randomly or systematically. The teacher then observes each student in terms of the behavior they are monitoring in the checklists to assess the progress of each student in acquiring the desired behavior, such as acquiring the skill of oral expression and good listening skills. The teacher usually decides the number of observations in which the observation is made without rating it each time. Checklists are considered a method of observation that focuses on analyzing learning outcomes. The teacher relies on determining the student's answer by asking multiple specific questions and setting an assessment scale, assigning a number of points for each answer, and then calculating the total score by adding up those points. (Maamria, 2012).

Example:

No	**Yes**	**Indicators**

6.1.3 Descriptive Grids: Each standard of the network is matched by a special descriptive scale with sequential steps.

Example :

	Scales	Standards
	Most rules are properly applied	Respect the Rules
	Few rules are properly applied	
	Does not apply any rule correctly	

6-2 Holistic Grids:

These grids consist of a single general scale with several levels. Each level of the grid contains a qualitative description of proficiency according to the levels of mastery.

3	The student clearly expresses his opinion and provides examples to support his opinion.
2	The student clearly expresses his opinion but does not provide examples to support his opinion.
1	The student does not express his opinion clearly, although he provides examples.
0	The student does not clearly express his opinion and does not provide examples to support his opinion.

The Interview

1- Definition of the Interview:

The interview is one of the assessment tools that employs the strategy of direct communication. It is a tool for collecting data through verbal interaction between individuals (Khattab, 2001, p. 94). It is also defined as a predetermined meeting between the student and their teacher, allowing the teacher the opportunity to obtain information regarding the student's ideas and attitudes towards a specific topic, and it includes a series of pre-prepared questions (Doumis, 2008, p. 92).

According to Mahmoud (2004), the interview is an educational measurement method and is a human interactive relationship between the teacher and the student, face-to-face, according to a scientific approach and suitable conditions of knowledge and experience. The teacher tries to understand the student's opinions under specific conditions of time and place. It is a tool for collecting and analyzing information to reach conclusions.

Thus, the interview seeks to gather information through communication activities about the progress made by the learner, as well as to understand their way of thinking and problem-solving style through questioning and answering. It may be prearranged, and the student is informed that it will occur either during or after the lesson, with its primary aim often being formative assessment.

In this context, the interview encompasses the collection of two types of information: the content acquired by the learner through following the presented lesson or lessons and the manner in which this content is approached through analysis, inference, deduction, and interpretation, along with all cognitive processes utilized, which are targeted in the study (Omar et al., 2010, p. 138).

The interview is a collaborative effort and an opportunity for the student to learn from the teacher or assessor, while the teacher gains insight into the student's thinking and attitudes in order to enhance the student's performance and improve the educational process.

2- Types of Interviews:

There are several types of interviews resulting from various classifications used to categorize interviews:

2-1 Based on the Degree of Specification:

According to the degree of specification, interviews can be categorized as follows (National Assessment Team, 2004, p. 65):

- Specific Interview:

In this type, questions are asked that require precise answers, leaving no room for additional questions.

- Non-specific Interview:

The questions do not have predefined answers, allowing for the addition of new questions based on the responses.

- Specific-Open Interview:

This type combines both specific and open-ended questions.

2-2 Based on the Number of Participants:

- Individual Interview:

It occurs between the assessor and the learner, aiming to gather information about the learner, including their abilities, preparedness, inclinations, and attitudes. Conducting an individual interview enhances credibility and grants the learner more freedom to express themselves without fear of making mistakes, as they can be corrected and adjusted by the assessor.

- Group Interview:

This involves meeting with a group of learners simultaneously in the same location. The interview's subject matter and the information to be collected or processed are common to all participants. It allows for more interaction, as learners benefit from each other and learn from correcting one another's mistakes, in addition to correcting their own.

2-3 Based on the Purpose:

- Research Interview:

Aimed at obtaining information, opinions, and understanding the factors and influences involved in the research subject. It enables participants to explore the meaning of questions and answers, facilitating the clarification of any misunderstandings promptly to ensure clear questions and answers, avoiding misinterpretation. (Mahmoud, 2004, p. 79).

- Diagnostic Interview:

Used to contribute to diagnosis, meaning it aims to identify problems, obstacles, and difficulties faced by the learner, understanding their causes, timing, and influencing variables, while identifying surrounding conditions and factors. Usually, the diagnostic interview is between a learner seeking assistance and an educator providing help. (Omar et al., 2010, p. 138) The diagnostic interview serves as a prelude to intervention and care, extending beyond mere diagnosis.

- Therapeutic Interview:

In these interviews, the assessor aims to address the learner's mistakes and to outline a plan to address the identified problems precisely.

3- Steps of the Interview:

The interview is an organized and purposeful process that involves essential steps, as outlined by Omar et al. (2010):

3-1 Goal Setting:

It is essential to define the objectives of the interview and translate them into criteria and indicators for observing their presence clearly and objectively. Moreover, determining the time, place, and creating suitable conditions for the interview are crucial.

3-2 Preparation of the Interview Guide:

This involves creating a form containing a list of organized questions, along with additional questions that can be used to clarify vague or incomplete answers, aiming to obtain comparable information. Additionally, specifying the method of presentation ensures the learner's understanding.

3-3 Conducting the Interview:

This step represents the actual execution of the interview, whether conducted individually or in groups. It involves ensuring comfort, tranquility, and confidence, fostering an atmosphere of warmth and trust. Constructing the interview relies on necessary cooperation to obtain responses and data reflecting the true level of the student, including their characteristics, attitudes, and inclinations.

3-4 Recording:

During the interview, events and information are recorded after confirming the accuracy of the students' responses. The evaluator records these responses on a pre-prepared form. It's worth noting that the information recorded may not necessarily be solely what the student verbalizes but may also include non-verbal cues and signals that the student may not intentionally convey. Questions can be used to assess the sincerity of the student's expression. Audio recording may be employed, but it must be disclosed to the student, although it may create an uncomfortable and unnatural environment for the student, thereby compromising the credibility of the gathered information.

3-5 Drawing Conclusions:

By collecting the recorded data and reviewing the interview, the evaluator can provide a final assessment based on the obtained information. This assessment involves thorough analysis and study, incorporating derived conclusions and the decisions that may result from them.

4- The Role of the Teacher and the Learner in the Interview:

The interview is a collaborative effort between the teacher and the learner or learners. The role of the teacher includes task definition, scheduling progress monitoring sessions, arranging interview sessions, preparing questions to guide learners, suggesting subsequent steps, facilitating understanding of the learner's perspective and justifications, delivering lessons to learners in areas where they face difficulty if necessary, and ensuring learner progress, as well as ensuring they have opportunities for interviews or meetings with their peers. On the other hand, the role of the learner involves positively engaging in the interview,

respecting time, being flexible in interactions, bearing responsibility, demonstrating learning, and expressing it. (National Assessment Team, 2004, p. 64)

5- Advantages of the Interview: (Abu Al-Futouh, p. 138)

The interview possesses several advantages that make it an effective tool in educational assessment. Some of these advantages include:

- Providing individuals with the opportunity to express themselves more freely compared to other assessment tools.
- Ease of gathering information from young children, especially those who have not reached a sufficient level of linguistic development, as appropriate verbal expression is facilitated.
- Flexibility and realism in gathering information due to the possibility of clarifying and explaining questions in case of difficulty for the respondent.
- Providing a more accurate picture of an individual's personality, indirectly inferred.
- Supplying information about personal, emotional, and psychological issues that cannot be directly expressed, especially in counseling or selection situations. (Khalil, 2011, p. 138).

6- Factors for Successful Interviews

The effectiveness of the interview as an assessment tool is linked to several factors. When these factors are present and carefully attended to, the results of the interview can be beneficial for improvement and modification, whether for the learner, the teacher, or the teaching process. Some of the factors mentioned by Alam (2015) include:

- Preparing in advance for the interview and ensuring suitable conditions.
- Ensuring that the student understands the purpose of the interview and how to use the information obtained.
- Carefully formulating questions directed at the student to be focused, clear, concise, and directly related to obtaining insights, feelings, and knowledge about the topics relevant to the interview's purpose.

- Encouraging the student to benefit from and respond to the interview, instilling confidence in them, and ensuring their understanding of the questions posed to them, as students from different cultural and social backgrounds may vary in their understanding of certain questions.
- Overcoming the problem of the student's responses being influenced by their perception of the teacher's authority, which may be a governing factor in the interview.
- Avoiding directing the student towards a specific direction or influencing them during an open-ended interview, especially to ensure they respond with what the teacher expects rather than the student's genuine answer.
- Actively listening to the student's responses to the questions posed to them.
- Preparing in advance for how to document the results of the interview, which may involve using an audio recording device or orally reading a questionnaire and taking comprehensive notes after the interview.
- Dealing with the assessment results in a way that serves the learner and the teaching process.

Portfolio of Achievement

1- Concept of the Portfolio:

The portfolio of achievement, or the portfolio, is a purposeful compilation of a student's work that illustrates their efforts, progress, and achievements in a specific academic field. This compilation should include the student's contributions in selecting the portfolio's content, as well as guidance in this selection, criteria for judging the quality of the work, and evidence of the student's self-reflection on these works (Alam, 2015, p. 157). Al-Barkati (2008) defined it as a repository for the best and most prominent achievements of the student in a course during a semester, highlighting and elucidating their progress and development during the course of study.

2- Objectives of the Achievement Portfolio:

Research, studies, and practical empirical experiments have proven that the achievement portfolio achieves several objectives, including those mentioned by Kojak (2012):

- To enable the student to actively participate in selecting and organizing the content of their studies.
- To acquaint the student with reflection on their work and engage in self-critique, learning objectivity and the ability to recognize their capabilities and judge their achievements.
- The portfolio aids in documenting what the student has learned in areas that are difficult to document through other assessment methods.
- The portfolio facilitates communication between the teacher and parents about the student's performance, strengths, and weaknesses.

3- Characteristics of the Achievement Portfolio:

The achievement portfolio possesses characteristics that contribute to the enhancement of the educational process, which are as follows:

3-1 Adaptability to Educational Diversity:

Each student prepares a separate portfolio containing their works and achievements, which the teacher then examines independently. Thus, the portfolio represents an educational outcome tailored to a specific student, aligning effectively with educational diversity.

3-2 Emphasis on Product Assessment:

It includes students' works or samples thereof, which are linked to curriculum outcomes. Additionally, it outlines the processes, procedures, and steps involved in the completion of the work.

3-3 Focus on Strengths:

Each student has strengths that can be enhanced. The portfolio emphasizes the students' strengths, encouraging them to deliver their best work. Therefore, they enjoy working on the portfolio as it caters to their preferences and aligns with their abilities.

3-4 Student Engagement in the Evaluation Process:

Both students and teachers develop the portfolio at all stages, from defining outcomes to completing and finalizing the works in the portfolio. The portfolio involves students in the evaluation process, which is a collaborative effort based on clear criteria. This gives students a sense of direct responsibility for their learning and provides them with a clear picture of what constitutes good work and achievement, guiding them toward the desired progress.

3-5 Facilitating Transfer and Exchange of Experiences:

Through the portfolio, the learner can showcase their abilities to others by providing tangible evidence of their learning through samples of work and achievements included in the portfolio. This provides a clear picture of their progress and development. By discussing the portfolio with the teacher and parents, opportunities for exchanging experiences through discussion, dialogue, and feedback are created.

3-6 Continuity:

The portfolio may require longer periods for evaluation, extending to more than a year. Additionally, reviewing the portfolio may require longer periods for periodic review by the teacher. Typically, the teacher first examines the

portfolio and then discusses it with the learner (National Assessment Team, 2004, p. 79).

4- Portfolio Contents:

There is almost consensus among scholars and educators on the elements of the portfolio as outlined by Al-Agha (2005).

4-1 Student Work Portfolio:

This documents some of the student's works, including the main ideas of the student and illustrating the efforts made in a specific subject area, such as writing a poem or critique, solving a mathematical problem, critiquing a book, or expressing an opinion on a particular topic.

4-2 Diverse Student Works:

This includes the various abilities and activities of the student in any field, such as linguistic, mathematical, or scientific fields, demonstrating the depth of the student's production in any area.

4-3 Student Reflections:

This involves the student's ability to express their opinion, describe the distinctive aspects of the work they have done, clarify the progress made as a result of this work, and the extent to which the student's work is linked to the intended goals.

Examples of portfolio contents as mentioned by Al-Abssi (2010, pp. 95-99):

- Samples of Student Writing: Portfolios are widely used in the field of language skills, especially writing, where they may include drafts of previous writings to track progress over time.
- Lists of Sources and Materials Used by the Student: Such as dictionaries, books, newspapers, journals, computer programs, and others.
- Self-Reflection Journals: Students can articulate their reflections on their work in a reflective journal by jotting down their thoughts and reflections on one part of the page, leaving the other part for the teacher to comment on these ideas. The teacher may also provide a list

of questions for the student to answer in their journal, such as: "What did you learn from the subject you studied today at school?"

- Student Reports: These include summaries of research papers and books, reports on laboratory experiments, observations conducted by students, interviews between students and their peers, teachers, or members of the local community, as well as reports on field trips.
- Group Activities: Developing social skills, social interactions, and cooperative learning are important aspects that schools should foster in students. Documenting these activities and including them in the portfolio can illustrate each student's contribution within the group framework.
- Test Scores: Portfolios typically include students' grades in achievement tests, especially if the purpose of the portfolio is to monitor the student's progress.

5- Categories of Achievement Portfolio:

The achievement portfolio (portfolio) has been classified into four categories, as mentioned by both Al-Kilani et al. (2014) and Al-Absi (2010).

5-1 Cumulative Portfolio:

In this type, the student records their work or achievements for selection or promotion purposes required for entry into a profession or program. This type of portfolio does not require feedback.

5-2 Training Student Portfolio:

This portfolio represents a sample of the student's work over a specified period. It gathers a collection of efforts required during the learning process, shedding light on the axis of knowledge and skills acquired.

5-3 Reflective Student Portfolio:

It consists of purposeful student work that provides evidence of progress and growth. By compiling these pieces of evidence, it showcases the best-selected practices that align with criteria consistent with self-progress over time.

5-4 Personal Developmental Student Portfolios:

It comprises a collection of student work aimed at personal evaluation of professional growth over a long-term process. This collection represents an opportunity for discussion and value provision for individual activities.

Self-Assessment and Peer Assessment

Assessment, as an educational process, aims at improvement and modification rather than just assigning grades to express achievement. In this context, there has been a growing interest in types of assessment that focus not only on grading but also on improvement and modification, including self-assessment and peer assessment.

1. Self-Assessment

1.1 Definition of Self-Assessment:

Al-Assaf (1995) defines self-assessment as the issuing of judgment by the student beneficiary of the course regarding the extent of their learning of the basic principles. In other words, self-assessment is when the student beneficiary of the course assesses themselves in terms of their ability to answer, understand, and master skills and problem-solving.

Self-assessment is the individual's assessment of themselves, based on the information they provide or the feelings they express about a situation, activity, or topic that interests or concerns them. The individual should keep a record in which they document their actions, activities, accomplishments, or decisions in a specific subject, their success or failure in performing the task. Then, they compare their current success in this task with what they achieved in a previous period. This comparison serves to motivate them to exert more effort to reach higher levels and can be represented graphically to facilitate comparison purposes between different periods. (Al-Sayed, 2011, p. 287)

Self-assessment goes beyond observation, analysis, and judgment of performance based on specific criteria, extending to the student's ability to develop plans to improve and enhance performance through mutual cooperation between the teacher and the learner. (National Assessment Team, 2004, p. 72)

Self-assessment is not limited to the learner but also extends to its use by the teacher, where the teacher reviews themselves in terms of lesson preparation, their interaction with students in classrooms, teaching strategies, exam adaptation, and the use of technology in assessment. For the learner to be able to self-assess, they must reach a sufficient level of maturity that allows them to perceive all aspects of this process, understand its objectives, and acquire the ability to assess their progress objectively, away from bias towards oneself. Encouraging them to engage in this process until they are convinced of it and enthusiastic about it, with sufficient training. (Tamam, Kamal, & Amin, 1997, p. 267)

1-2 Advantages of Self-Assessment:

Self-assessment offers several advantages, among which are the following:

- It helps the learner discover their mistakes, enabling them to rectify them promptly, thereby preventing the acquisition of erroneous habits that would be challenging to change later on.

- The learner's recognition of their own mistakes and weaknesses makes them more receptive to criticism from others because they critique themselves first. Consequently, they are less affected when criticized by others.

- The process of self-assessment makes the learner more tolerant of others' mistakes, as their experience in this area has taught them that every learner has their own shortcomings.

- Self-assessment teaches the learner to take responsibility and instills self-confidence. (Tamam, Kamal, & Amin, 1997, p. 267)

- Learning self-assessment does not stop at formal education but is an ongoing process for every individual, aimed at adding to their knowledge, skills, and abilities in judgment and action, adapting them to themselves and their environment in terms of knowledge, work methods, living with others, and life skills. (Shahata, 2012, p. 356)

- Self-assessment aligns with the principles of constructivist cognitive learning, where knowledge is built, and the learner is the one who learns and constructs knowledge.

- It reduces the stress and anxiety typically associated with assessments. (Zeitoun, 2007, p. 648)

1-3 Methods and Tools of Self-Assessment:

Zeitoun (2007) outlined some methods and tools for self-assessment, including:

1-3-1 Self-Review and Self-Evaluation Checklist:

Utilizing a self-review and self-evaluation checklist is beneficial in assisting students to assess their work and accomplishments. Both teachers and students can collaborate and participate in developing these checklists in line with the objectives and goals, ensuring concerted efforts that encompass the elements and components to be evaluated in processes and outcomes. Subsequently, the student places a mark to indicate the extent to which performance criteria and standards have been met.

1-3-2 Questionnaires:

Students can use questionnaires to review some tasks and procedures they undertake to ensure the quality of their performance. These questionnaires cover a wide range of elements or components based on the processes and outcomes to be evaluated, with the student autonomously determining their level of attainment or achievement of these elements.

1-3-3 Self-Testing:

This can be done through taking tests that were administered previously or through tests similar to those given in the classroom. The student then refers to the answers provided in the reference material or to the model answers prepared by the teacher who designed the test in light of previous objectives, goals, and criteria.

1-3-4 Self-Evaluations:

In this method, students use credible tools to draw a self-image, including personal characteristics, cognitive preferences, learning styles, vocational inclinations, scientific interests, and orientations. The student identifies their strengths to reinforce them and their weaknesses to address them. These tools are available at the school level or on the internet, easily accessible to students.

1-3-5 Reflective Questions:

Students can express their reflections and insights on what they have learned or are learning. These questions focus on learning, and teachers may assist students in improving their work by teaching them how to refine and enhance it when they engage in reflective activities.

2. Peer Assessment

2-1 Definition of Peer Assessment:

It is known as the assessment conducted by a group for one of its members or the assessment by an individual of the members of the group to which they belong. This method takes various forms, including:

- The teacher individually asks students for their opinions about their classmates, then carefully and accurately assesses this information.

- The teacher records specific attributes and asks the student to write the names of their peers who exhibit each of these attributes.

- The teacher divides the students into groups and asks each student to write a brief report covering the strengths and weaknesses of each student in the group to which they belong. (Al-Sayed, 2011: 288).

Peer assessment and its advantages integrate with self-assessment and become evident in light of the following points:

Learners become more systematically involved in the assessment process.

Peer assessment develops skills and habits of constructive criticism (analysis, classification, inference, and evaluation) both currently and in the future within real-life professional settings.

It provides an environment conducive to giving learners the opportunity to establish performance standards and criteria in the assessment process.

It enhances learners' self-confidence and motivates them to take responsibility for their learning, similar to self-assessment.

It contributes to the development of personal and social skills, improves the quality and fairness of teaching and assessment, and may prevent entry into biased assessments, as learners are evaluated by their peers.

It helps learners recognize the good work they have done by evaluating it, leading to a better understanding of the subject matter and reducing psychological exam stress.

It fosters collaborative efforts among students to develop performance standards and criteria that can be relied upon in the assessment process and judgment issuing.

It provides the opportunity for peer feedback (beyond the teacher's authority), review, improvement, and enhancement of works and learning, allowing learners to play an active and positive role in their learning and self-assessment through active participation in cognitive, performance, and evaluative tasks.

Peer assessment contributes to the development of certain personal qualities and values, such as respecting others' opinions, appreciating ideas, objectivity, open-mindedness, and self-respect and appreciation. (Zeitoun, 2007, p. 653).

2-3 Peer Assessment Methods:

Several methods and practices can be followed for peer assessment, including:

1- Presenting individual works and achievements to peers in a discussion session or peer seminar where their opinions, ideas, and criticisms are evaluated against predefined performance criteria and standards. This provides learners with feedback to review, refine, and improve their work, enhancing performance, learning, and refinement.

2- Presenting group works where each small cooperative group presents their works and achievements to peers or other peer groups. Peers then discuss these works, activities, experiments, and evidence in light of performance criteria. Based on this, the group can review, refine, and enhance their works, strengthening strengths and addressing weaknesses to improve learning and achieve goals.

3- Due to the close relationship between peer assessment and self-assessment in the assessment process, the same methods and techniques mentioned in self-assessment can be used, such as review and estimation lists, peer assessment questionnaires, and their forms, in light of performance criteria and standards. (Zaytoon, 2007, p. 654)

4- By allowing students to choose a peer or peers to evaluate their accomplishments and works against specific criteria or collaborate on setting them if they are not already established.

Assessment of Aptitudes, Abilities, Interests, and Attitudes

1- Aptitude Assessment:

Aptitude assessment is an evaluative process aimed at understanding students' level of preparedness before starting the learning process. This assessment is a critical step to ensure that teaching is adapted to meet students' needs and capabilities. This type of assessment enhances teaching effectiveness and improves educational quality by identifying students' strengths and weaknesses.

1-1 Definition of Aptitudes:

The aptitudes are individual possessions, which enable them to perform specific tasks efficiently and effectively. These aptitudes are not limited to skills acquired through direct learning but also include latent abilities that may not have been fully developed yet. Aptitudes are significant indicators of potential success in specific areas, such as academic achievement, professional performance, or creative activities. (maamria, 2012)

For example, aptitudes include general intelligence, linguistic abilities, logical reasoning, creativity, and motor skills. Psychological measurement tools, such as tests and questionnaires, are used to assess these aptitudes and determine an individual's level of ability in various domains.

1-2 Importance of Aptitude Assessment:

The importance of aptitude assessment lies in its ability to provide clear insights into students' prior knowledge and foundational skills. According to a review of studies by Peter Black and David Wiliam (1998), aptitude assessment helps identify gaps in knowledge and skills, allowing teachers to design instructional strategies that cater to each student’s needs. This type of assessment enables teachers to provide necessary support to students before introducing new content, thus enhancing students' success in subsequent educational stages.

1-3 Methods of Aptitude Assessment:

Aptitude assessment involves several tools and techniques, including:

- **Diagnostic Tests**: These tests are used to evaluate students' knowledge and skills before starting a specific topic. They help establish a foundation upon which new knowledge can be built.
- **Surveys and Interviews**: These are used to gather information about students' backgrounds and prior experiences.
- **Direct Observation**: Teachers can use direct observation to understand how students handle specific tasks and identify their strengths and weaknesses.

1-4 Applications of Aptitude Assessment:

Implementing Aptitudes assessment allows for the design of customized instruction that meets individual students' needs. Based on the results of this assessment, teachers can tailor the educational content and adjust teaching methods to ensure all students start on an equal footing. This customization enhances learning effectiveness and reduces gaps between students.

2- Ability Assessment:

Ability assessment is a process aimed at measuring individuals' capacity to perform specific tasks or achieve defined goals based on their skills and knowledge. Ability assessment is a fundamental tool in education and training, as it helps in determining students' progress and guiding them toward achieving their educational objectives.

2-1 Definition of Ability: Abilities are the fundamental potentials that an individual possesses, allowing them to perform specific tasks effectively and efficiently. Abilities are essential factors that impact an individual's learning, development of new skills, and execution of various tasks. Abilities vary from cognitive, motor, and social to creative abilities (Omar et al., 2010, p. 307).

2-2 Differentiating Between Aptitudes and Abilities:

Aptitudes refer to the potential and capacity to acquire and learn specific skills. They represent the natural inclination or tendency an individual has towards learning new skills or performing certain tasks.

Abilities refer to the actual skills and capabilities that an individual possesses, which enable them to perform specific tasks efficiently at the present time.

Interaction between aptitudes and abilities: Aptitudes can guide an individual towards developing certain abilities. For example, a person's aptitude in the artistic field may lead to the development of advanced artistic skills. At the same time, abilities can enhance or support aptitudes. For instance, if an individual has a strong ability for quick learning, it may enhance their aptitude for developing new skills in different areas.

This interaction between aptitudes and abilities helps in understanding how individuals can grow and develop in specific fields based on their natural inclinations and current capabilities.

2-3 Importance of Ability Assessment:

The importance of ability assessment lies in its ability to provide accurate insights into individuals' skills and capabilities. According to John Hattie's study on the effects of assessment (2009), ability assessment plays a central role in improving the effectiveness of teaching and learning. Through this assessment, teachers and trainers can identify individuals' strengths and weaknesses, helping to design suitable educational and training strategies. Ability assessment also aids in guiding individuals toward achieving their goals and improving their performance through constructive feedback.

2-4 Methods of Ability Assessment:

Methods of ability assessment include a variety of tools and techniques, such as:

- Practical Tests: These measure individuals' ability to perform specific tasks based on their practical skills. For example, tests that assess problem-solving abilities or interaction with practical scenarios.
- Self-Assessments: Individuals are asked to evaluate their own skills and abilities. These assessments enhance self-awareness and help individuals identify areas for improvement.
- Project-Based Assessments: Involve assigning individuals tasks or projects that evaluate their ability to apply skills and knowledge in practical contexts.

2-5 Applications of Ability Assessment

Ability assessment is an effective tool for determining individuals' suitability for various roles or jobs. In education, it helps guide students toward areas where they excel and need to enhance their abilities. In training and employment, it can be used to select individuals most suitable for positions based on their ability to perform required tasks.

3- Assessing Interests and Attitudes:

Assessing interests and attitudes is a process aimed at understanding individuals' preferences and orientations toward certain fields of study or work. This assessment reflects the degree to which individuals prefer specific activities and their future goals, aiding them in making informed decisions regarding their educational and career paths. This process is essential in educational and vocational guidance as it helps individuals achieve their aspirations and satisfaction.

3-1 Definition of Interests and Attitudes:

Interests are an individual's preferences, which may be specialized, professional, or personal in nature. Interests can be related to the profession that the individual practices, educational interests related to the subjects and courses they study, or hobby-related interests that the person engages in, such as various hobbies. It is noteworthy that these interests maintain a relative degree of stability over time. (Omar et al., 2010, p. 307)

Alport (1954) defines an attitude as a mental and neural readiness organized by personal experiences that directs an individual's responses to objects and situations related to this readiness. An attitude is a state that drives an individual to either support or oppose a subject or action. It represents preferred or non-preferred responses and reactions towards stimuli, whether they are individuals, groups of people, social norms, or institutions. Attitudes consist of three components: cognitive, behavioral, and emotional. (Omar et al., 2010, p. 318).

3-2 The Importance of Assessing Interests and Attitudes

Assessing interests and attitudes provides valuable insights into individuals' preferences and orientations. According to John L. Holland (1997), this assessment helps individuals identify their personal inclinations and

orientations toward specific fields, contributing to the selection of appropriate educational and career paths. This assessment enhances personal and professional satisfaction while reducing the likelihood of frustration or mismatch in the future.

Researcher Richard Katz (2004) also points out that assessing interests and attitudes can help understand the psychological factors influencing individuals' decisions. This assessment can identify the factors that affect decision-making in education and work, thereby improving the effectiveness of guidance and counseling.

3-3 Methods of Assessing Interests and Attitudes:

The methods of assessing interests and attitudes include a variety of tools and techniques, such as:

- **Interest and Attitude Surveys**: Used to explore individuals' preferences and orientations toward certain fields. For example, the "Holland Vocational Interest Inventory" can be used to identify individuals' inclinations in various work fields (Holland, 1997).
- **Personal Interviews**: Involve discussing individuals' interests and orientations, helping to analyze their inclinations based on dialogue and personal interaction.
- **Psychological Assessments**: Include tests that measure personality traits and attitudes, such as personality tests that help understand individuals' interests and orientations toward specific fields (Katz, 2004).

3-4 Applications of Assessing Interests and Attitudes

The results of assessing interests and attitudes assist in guiding individuals toward fields that align with their interests and abilities. In education, this assessment can help students choose majors that match their interests and orientations, enhancing their academic success. In the workplace, interest and attitude assessments can help in selecting suitable careers that correspond with individuals' preferences, increasing their job satisfaction and performance.

Electronic Assessment and Question Banks

Scientific and technological advancements have led to the emergence of electronic education, whether independently or integrated with traditional education. This has led to the use of tools to verify learning and its effectiveness through effective electronic assessment, which aids in the development and diversification of educational processes and scientific fields. This electronic assessment can also be applied to traditional education in various forms.

1- Definition of Electronic Assessment:

Ismail (2009) defines electronic assessment as the process of employing information networks, computer equipment, educational software, and multimedia educational materials using assessment methods to collect and analyze student responses. This assists faculty members in discussing and determining the effects of programs and activities on the educational process to reach a judgment based on quantitative or qualitative data related to academic achievement.

From this definition, all types of assessment can be fully applied electronically, albeit with some relativism. In e-learning, assessment must adhere to the science of learning through diagnostic assessment before instruction begins, formative assessment to gauge student progress and provide feedback, and summative assessment through electronic testing to determine success. The integration of technology into assessment has provided numerous possibilities that cannot be achieved through traditional assessment methods.

2. Methods and Tools of Electronic Assessment:

2-1 Surveys and Questionnaires:

In this method, students are asked to respond to a questionnaire regarding electronic learning programs, yielding results that are largely credible. Surveys

are frequently used in assessing conferences and meetings, where a questionnaire is designed and distributed electronically, such as using Google Forms to build survey models.

2-2 Personal Interviews:

Through personal interviews, we can assess the effectiveness of the program based on student responses. Interviews can be conducted face-to-face via camera, audio, in writing, or in a combination of these methods.

2-3 Observation and Application:

In this method, students are placed in practical applications and practice scenarios, where their progress in skills is observed during the activity using observation cards. This observation can be either direct or indirect, through recordings, product demonstrations, or completed projects.

2-4 Electronic Achievement Tests:

These are tests similar to traditional paper-based tests but administered via computer, allowing remote answering. They provide a wider range of questions that are difficult to implement on paper. These tests may be associated with e-learning or blended learning, which combines both in-person and online learning, a widely adopted approach. Many programs are available for creating electronic tests independently or as add-ons in distance learning platforms like Google Classroom.

2-5 Electronic Portfolios:

The widespread use of the portfolio tool in assessment has led to its electronic utilization, which enhances its components because it can include videos or programs that cannot be converted into paper format. Additionally, it offers ease of transfer and assessment (Alam, 2015, p. 163).

2-6 Self-Assessment:

We can benefit from electronic assessment by implementing self-assessment, where, after reviewing lessons or parts thereof, students test themselves by answering prepared questions. They can then review the answers and grades, serving as feedback to gauge their progress.

3- Characteristics of Electronic Assessment:

Electronic assessment has several features and advantages. Abdelhameed (2005) outlines some of these characteristics:

- Interactivity:

Refers to presenting tasks to students and the ability to provide quick feedback on their actions. It involves interaction between the learner and what the computer presents, allowing learners to control, adjust, and choose between available options.

- Synchronous Interaction with Diverse Students:

Different interactions with diverse students can occur simultaneously.

-Variety and Breadth of Tools:

Assessment tasks can be presented through multiple media, making tasks more realistic and diverse and presented in various ways.

- Network Utilization:

Refers to the interconnectedness where institutions, schools, parents, government officials, authors, test reviewers, graders, and students are electronically connected through communication networks.

- Standardization:

Means that the network operates according to a set of predefined rules, allowing for easy exchange of information and entry into various computer environments. The internet enables the presentation of any content to a large number of people, obtaining data instantly, processing this data, and making information available anywhere in the world at any time.

4. Benefits of Electronic Assessment:

• Flexibility and Time Saving:

Tests can be created, edited, and reused as needed. Tests can be distributed and answers obtained via the internet, saving preparation and distribution time.

• Reduction of Feedback Time:

Reduces the time needed for students to receive their results and feedback, allowing them to use the knowledge gained to address deficiencies as quickly as possible.

• Reduction of Required Resources:

Human resources can be minimized by electronically grading simple tests. Programs can reduce processing time and grading assignments. Assignments can be completed, collected, graded, and commented on entirely electronically.

• Record Keeping:

Each student's record can be automatically maintained, and these records can be stored centrally for relevant parties to access.

• Facilitation:

Achieving comfort and ease for all parties, thus facilitating their adaptation. Tests are automatically collected and graded, meaning students can receive immediate feedback and submit assignments and performances at their convenience.

• Ease of Data Utilization:

Because assignment grading data is stored electronically, it is easier to analyze and utilize (Abdelhameed, 2005, p. 226).

5. Challenges of Electronic Assessment:

- Security Concerns: The use of information networks and remote testing can make questions, student answers, and results vulnerable to hacking, requiring extensive software and procedures to address this issue.
- Possibility of Cheating: Students may share exam questions and answers in real-time, but this can be mitigated through controlled timing, randomized question order, and alternative options.
- Cheating from Unauthorized Sources: Students may access prohibited websites, use crib notes, or even rely on paper-based sources for cheating.
- Impersonation: Impersonation: Another person may answer on behalf of the student, assuming their identity.
- Technical Disruptions: Devices and software may malfunction during exams, as they are typically timed. Additionally, internet disconnections or low bandwidth can disrupt the assessment process (Khalil, 2017, p. 36).

6. Stages of Designing and Producing Electronic Tests:

6-1 Analysis Stage:

In this stage, the overall test objective, test taker characteristics, educational material analysis, and real-world analysis are determined.

6-2 Design Stage:

This involves writing test questions in various formats, such as objective or essay questions, specifying test instructions, duration, and multimedia used; feedback methods; determining the model answer; and correction methods—whether automated or manual, requiring the intervention of the grader. It's worth noting that there is software available for this purpose.

6-3 Publication and Electronic Distribution Stage:

The test is published online, on discs, or on digital media, and distributed to students in their locations.

6-4 Application Phase:

Implementing the test on the relevant students, gathering data through students' responses, and conducting automated or manual corrections, followed by announcing the results of the electronic test.

6-5 Evaluation Phase:

Assessing the validity of the electronic environment, its transportation validity, and its delivery in terms of ensuring test confidentiality.

7- Question Banks:

7-1 Definition of Question Banks:

Randor defines them as a collection of files stored on a computer system containing suitable types of test items. These items are classified and encoded according to educational specifications, academic levels, and measurable educational objectives. They possess various distinctive features such as difficulty and discrimination levels. These files are stored within a specified database, referred to as question banks. They are specialized electronic programs designed to store a vast number of high-quality questions in terms of difficulty and discrimination. These questions are indexed and categorized for easy retrieval in test preparation.

7-2 Advantages of Question Banks:

Khalil (2017) enumerates several advantages of question banks, including:

• Saving teachers' time and effort, thus reducing financial costs.

• Minimal or no errors in question texts due to their prior statistical analysis and thorough examination.

• Ease of adding or removing items, facilitating quick access to tests in a short period.

• Accuracy in measuring students' achievement using high-quality questions.

• Ease of test administration in remote locations and with large numbers of test-takers using communication networks.

• Easy use of printed tests if paper-based administration is preferred.

• Ability to shuffle question order and randomize items to prevent cheating easily.

7-3 Special Software for Question Banks:

Technological advancements have led to the emergence of various software programs dedicated to building question banks and using them in test design. Each program distinguishes itself from others with features such as ease of use, question security, and the ability to incorporate images and videos. Some notable programs include Fast Test, Fast Test Professional, The Examiner, Exam Manager, and BILOG-MG (Kheder, 2007, p. 128). Additionally, there are many other programs available, each serving similar purposes. With these software programs, users can create numerous question banks tailored to specific subjects or disciplines. These question banks can then be easily accessed and utilized to generate tests as needed. Through their storage capabilities, these programs greatly facilitate the gradual accumulation of rich question banks over the years.

Educational Assessment Challenges

Through our study of the concept of assessment, its various types, tools, and how we can apply them in the field, especially with the evolving role of assessment in the educational process, the development of assessment methods has resulted in some problems. Let's review some of them:

1. Assessment Itself:

The presence of assessment in education and the decisions and judgments made based on it significantly impact students' educational trajectories. There has been a shift towards focusing more on assessment than on learning itself. Acquiring high grades has become more important, even if they do not reflect the extent of learning. For example, focusing on what is known as the "threshold" of lessons or determining which lessons will be included in the assessment can lead students to neglect lessons that are not included. The value placed on assessment, especially exams, has created fear and anxiety in students before, during, and after exams, known as exam anxiety, which is one of the psychological problems associated with assessment. Additionally, the grades obtained in education often determine a person's overall character, not just their academic performance, leading those with low grades to feel low self-esteem. Moreover, societal perceptions of them as failures can significantly impact them and often lead to deeper problems such as dropping out of school, which is a fertile environment for crime and deviance.

2. Overreliance on a Single Method and Strategy for Assessment:

Relying solely on achievement tests as an assessment tool, which is widely used, has drawn many criticisms. The popularity of achievement tests is attributed to their ease of preparation and low cost. However, there are many criticisms directed at the overreliance on tests as the sole tool in the educational

assessment process. Tests are considered ineffective in measuring complex skills for problem-solving, as well as in communication and critical thinking. They do not assess learners' personal thinking and mostly focus on the cognitive aspect, neglecting other essential aspects of personality formation, such as personal skills and the extent of their connection to the external environment. Additionally, realistic assessment, which assessment aims for, cannot be achieved through tests alone.

3. Poor Construction of Assessment Methods:

The success of the assessment process, in all its forms, depends primarily on the tools used. One of the problems encountered is the poor construction of assessment tools. Often, the specified assessment objectives and established criteria differ from the content of the tools. This discrepancy may aris due to the complexity of the human characteristics being assessed. The learning outcomes intended for assessment may be abstract and general rather than tangible and specific. It can be challenging to define and express them clearly and precisely.

Furthermore, the lack of suitable and reliable tools, coupled with technical inadequacies, whether from the tool developers or the environment in which they are used, poses numerous challenges in using electronic assessment or preparing question banks and effectively utilizing them to create quality tests.

4. Detachment of Assessment from the Educational Process:

The clarity of assessment objectives and their alignment with educational goals ensures that teaching and assessment are aligned. However, one of the problems is that assessment often adopts different and sometimes conflicting objectives from those of the course or subject matter, affecting the educational process and the performance of both teachers and students, especially when the test or assessment tool is not developed by the teacher.

Additionally, the presence of secondary objectives within the educational system that are entirely different from the assessment objectives, such as automatic promotion and increasing the pass rate, without directing or

transferring learners to more suitable studies or training, can further complicate the educational landscape.

5. Correcting Examinations and the Pitfalls of Subjectivity in Assessment:

Subjectivity and the lack of objectivity in grading exams, especially regular ones graded by teachers, make it difficult to accurately determine students' true levels or to compare results among students when there is variation between teachers or schools. The errors that assessors may fall into during grading and assigning grades are outlined by Khattab (2001) as follows:

5-1 Halo Effect:

This is the tendency of the assessor to let the initial general impression of the person being assessed influence their evaluations across all exams. This impression leaves a general direction and feeling towards the person, affecting grading or score allocation. Assessors may lean towards giving high grades to hardworking individuals and low grades to those perceived as less diligent.

5-2 Central Tendency Error:

Assessors tend to avoid extreme evaluations, neither giving very high nor very low grades.

5-3 Leniency Error:

This is the inclination to give average or higher grades to all individuals, avoiding giving low grades or grades below average.

5-4 Strictness Error:

Conversely, this is the inclination to give average or low grades to all individuals and not award high evaluations.

5-5 Stereotypical Error:

This is the assessor's tendency to give high grades to a particular ethnic, religious, or professional group.

5-6 Logical Error:

This is the personal bias in the relationship between assessments, such as leaning towards giving high grades in a specific subject because the assessor knows that the student has achieved high results in another subject perceived by the assessor to be related.

6- Inadequate Competence in Assessment:

One of the problems in assessment is the lack of competence among those involved in the assessment process. Often, this is not due to their personal capabilities but rather to the lack of training and development. Many find themselves tasked with preparing tests, observation grids, or any other assessment tool without receiving any prior training or development, especially since many assessors have not received basic training before serving in this role. Additionally, the evolution of assessment and the adoption of new methods and approaches have not been accompanied by continuous training to keep up. Technical weaknesses and lack of control over technology, whether in terms of hardware, software, or network connectivity, have all contributed to numerous challenges that hinder assessment from fulfilling its role in the educational process.

7- Absence of Providing Feedback:

Feedback is a fundamental element in our definition of assessment, in addition to measurement and estimation. Focusing solely on measurement and being content with assigning grades to students deprives assessment of its fundamental meaning. One of the problems in assessment is the failure to provide students with feedback and the complexity of its pedagogical processing. Limiting the concept of assessment to evaluation only and focusing on summative assessment while neglecting formative assessment, even considering summative assessment as a formative one, exacerbates this issue.

8- Difficulty in Comprehensive Assessment:

Comprehensive assessment is crucial in evaluating the educational system, curricula, and students' performance. However, one of the problems is the absence of this type of assessment, relying instead on partial analytical assessment. This problem may be due to the lack of mechanisms that enable comprehensive assessment, or their complexity if they exist. Additionally, the absence of specific standards to refer to and the lack of regulated tools

contribute to this issue, as well as the absence of a unified database to extract this type of assessment.

REFERENCES

- Abbas, Faisal. (1996). Psychological Tests: Techniques and Procedures. Beirut, Lebanon: Dar Al-Fikr Al-Arabi.
- Abdelhadi, Nabil. (2001). Educational Measurement and Assessment and Its Use in Classroom Teaching. 2nd ed. Jordan, Amman: Dar Wael for Publishing.
- Abdelhameed, Mohamed. (2005). E-Learning System. Cairo: Alam Al-Kotob.
- Abu Anab, Saad. (2000). Approaches and Trends in Educational Assessment.
- Abu Hawaj, Marwan et al. (2002). Measurement and Assessment in Education and Psychology. Amman: Dar Al-Ilmiyya International for Publishing and Distribution.
- Abu Jawad, Saleh Mohammad Ali. (2005). Educational Psychology. 4th ed. Amman: Al-Maseera Publishing, Distribution, and Printing House, Jordan, p. 411.
- Al-Abssi, Mohamed. (2010). Realistic Assessment in the Teaching Process. 1st ed. Oman: Dar Al-Maseera for Publishing and Distribution.
- Al-Agha, Abdel Mawla. (2005). Work Portfolios: An Introduction to Contemporary Assessment Approaches. Islamic University Journal, Volume 13, Issue 1, pp. 123-138.
- Alam, Salahuddin Mahmoud. (2002). Educational and Psychological Measurement and Evaluation. Cairo: Dar Al-Fikr Al-Arabi.
- Alam, Salahuddin Mahmoud. (2006). Psychological and Educational Tests and Scales. 1st ed. Jordan, Amman: Dar Al-Fikr.
- Alam, Salahuddin Mahmoud. (2015). Educational Measurement and Evaluation in the Teaching Process. 5th ed. Jordan, Amman: Dar Al-Maseera.
- Al-Assaf, Saleh bin Hamad. (1995). Self-Assessment for Researchers in Behavioral Sciences. 2nd ed. Riyadh: Al-Obeikan Library.

- Al-Atoum, Adnan Youssef. (2005). Educational Psychology: Theory and Application. 1st ed. Jordan: Dar Al-Maseera for Publishing, Printing, and Distribution.
- Al-Barkati, Niveen. (2008). The Effectiveness of Using Achievement Files on the Performance of Female Mathematics Students in the Activity Design Course in the Educational Preparation Program at Umm Al-Qura University.
- Al-Eissawi, Abdul Rahman. (1996). Measurement and Experimentation in Psychology and Education. Alexandria, Egypt: Dar Al-Ma'arif Al-Jameya.
- Al-Faqih, Ismail Mohamed. (2005). Psychological and Educational Evaluation and Measurement. Egypt: Dar Al-Ghareeb for Printing, Publishing, and Distribution.
- Al-Fayruzabadi, Majd Al-Din. (2008). Al-Qamoos Al-Muhit (The Comprehensive Dictionary). Edited by Al-Shami, Anas Mohamed, and Ahmed Zakaria Gaber. Cairo: Dar Al-Hadith.
- Al-Harthy, Salah Raddad (Ed.). Continuous Assessment from Theory to Practice. Saudi Ministry of Education.
- Al-Kilani et al. (2014). Measurement and Evaluation in Learning and Teaching. Open University of Jerusalem, Oman.
- Al-Muhaimid, Saadah; Al-Siriy, Sheikha; Al-Ajlan, Aziza. (2014). Some Models of Evaluation in the West (Taylor, Scriven, Stufflebeam, Stake, Hammond, Rogers). Retrieved from: http://master1435-4.blogspot.com/2014/03/blog-post_16.html
- Al-Nour, Ahmed Yakoub. (2007). Measurement and Evaluation in Education and Psychology. Al-Janadriyah for Publishing and Distribution, Jordan.
- Al-Rawashdeh, Ibrahim et al. (2000). Teacher's Guide to Building Achievement Tests. Ministry of Education, General Directorate of Examinations, Oman.

- Al-Sayed, Ali Mohammed. (2011). Encyclopedia of Educational Terminology. 1st ed. Oman: Dar Al-Maseera for Publishing and Distribution.
- Al-Shammari, Abdul Karim. (2013). Assessment: Types, Objectives, Models. Retrieved from: http://abdulkrem556.blogspot.com/2013/05/blog-post.html
- Al-Zaghloul, Emad Abdel Rahim. (2009). Principles of Educational Psychology. 2nd ed. Dar Al-Maseera for Publishing and Distribution, Jordan.
- Aziza, Rahma. (2016). Your Guide to Crafting Multiple-Choice Questions Tests. Damascus: Center for Measurement and Evaluation in Higher Education.
- Biggs, Kelly. (2014). Formative Assessment in Science: 75 Practical Strategies for Linking Assessment, Teaching, and Learning. Translated by Jaber Bin Mohammed Al-Jabr. Saudi Arabia: King Saud University Press.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7-74
- Bourguiba, Dawood. (2005). The Ideal Education Curriculum. Ghardaia, Algeria: Arab Printing Press.
- Bro, Mohammed. (2010). The Impact of School Guidance on Academic Achievement in Secondary Education. Dar Al-Amal, Algeria.
- Doumis, Mustafa Nimr. (2008). Modern Educational Assessment Strategies and Tools. Amman: Dar Ghayda for Publishing and Distribution.
- Ghanem, Mahmoud Mohamed. (2002). Educational Psychology. 1st ed. Jordan: Dar Al-Ilmiyyah Al-Dawliyyah for Publishing, Distribution, and Dar Al-Thaqafah for Publishing and Distribution.
- Hanouda, Ali. (2013). Social Interaction and Its Relationship to Academic Achievement Among Some Secondary School Students: The Final Classes at Badi Maki Bziriba High School in Zeribet El Oued -

Biskra - as a Model. Master's thesis, Faculty of Humanities and Social Sciences, University of Mohamed Khider Biskra.

- Hattie, J. (2009). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Routledge.
- Holland, J. L. (1997). Making Vocational Choices: A Theory of Vocational Personalities and Work Environments. Psychological Assessment Resources.
- Ibrahim, Magdi Aziz. (2008). Contemporary Educational and Teaching Issues. 1st ed. Cairo: Nahdet El-Sharq Printing, Publishing, and Distribution House.
- Ismail, Al-Ghareeb Zaher. (2009). Electronic Curricula: Design, Production, Publication, Implementation, and Evaluation. Cairo: Dar Al-Fikr Al-Arabi.
- Karagol, Ibrahim & Adiguzel, Oktay Cem. (2023). "Hammond's Evaluation Model" Journal ODUSOBIAD ODU SOsyal BIlimler Arastırmaları Dergisi 13 (1), 1433-1448
- Katz, R. (2004). The Motivation to Work: A Psychological Analysis. Routledge.
- Khalil, Hanan Hassan. (2017). Electronic Assessment. Oman: Al-Maseera Publishing, Distribution, and Printing House.
- Khalil, Mohammed Abu Al-Fattouh. (2011). Educational Evaluation: Between Reality and Aspiration. Riyadh: Al-Shaqra Library for Publishing and Distribution.
- Khattab, Ali Maher. (2001). Measurement and Evaluation in Psychological, Educational, and Social Sciences. 2nd ed. Egypt: Anglo-Egyptian Library.
- Kheder, Adel. (2007). Question Banks Between Theory and Application. Cairo: Dar Al-Sahab.
- Kojak, Kauther. (2012). Portfolio in Education: A Comprehensive Vision. Alam Al-Kotob, Cairo.

- Maamria, Bashir. (2007). Psychological Measurement and Tool Design. Ink Publications, Algeria.
- Maamria, Bashir. (2012). Fundamentals of Psychological Measurement and Tool Design. Batna: University of El Hadj Lakhdar, Dar Al-Khaldoonia for Publishing and Distribution.
- Mahmoud, Hamdi Shaker. (2004). Educational Evaluation for Teachers. 1st ed. Hail: Dar Al-Andalus.
- Makadem, Abdelhafidh. (1993). Statistics and Psychological and Educational Measurement. Algeria: University Publications Council.
- Malham, Sami Mohammed. (2001). Psychology of Learning and Teaching. 1st ed. Jordan: Dar Al-Maseera for Publishing, Distribution, and Printing.
- Mikhail, Imtanius Naeif. (2015). Psychological and Educational Measurement and Evaluation for the Disabled and Special Needs. 1st ed. Oman: Dar Al-I'asar Al-Ilmi.
- Moulay, Boudkhili Mohamed. (2004). Different Methods of Motivation and Their Relationship to Academic Achievement. University Publications Council, Algeria.
- Murad, Salah Ahmed, & Amin, Ali Suleiman. (2005). Tests and Measures in Psychological and Educational Sciences. 2nd ed. Dar Al-Kitab Al-Hadith.
- National Assessment Team. (2004). Evaluation Strategies and Tools. Jordan Testing Directorate.
- National Curriculum Committee. (2016a). General Framework for Middle School Education Curriculum. Algeria: Ministry of National Education.
- National Curriculum Committee. (2016b). Introduction to the Document Accompanying Middle School Education. Algeria: Ministry of National Education.
- Omar, Mahmoud Ahmed; Fakhro, Hessa Abdulrahman; Al-Subaie, Turki; & Turki, Amina Abdullah (2010). Psychological and Educational

Measurement. Dar Al-Maseera for Publishing, Distribution, and Printing. Amman, Jordan.

- Qattami, Abdellatif. (2014). Complex Thinking and Its Relationship to Academic Achievement: A Field Study on Middle Schools in Al-Wadi. Master's thesis, University of El Agouat.
- Qattami, Abdellatif. (2015). The Use of Specification Tables by Middle School Teachers in Constructing Achievement Tests: A Field Study in Al-Wadi. Journal of Seraj in Education and Social Issues, Faculty of Social and Human Sciences, University of Al-Wadi, Issue 1, pp. 170-185.
- Qureshi, Mohammed. (2002). Anxiety and Its Relationship to Academic Adjustment and Achievement Among Secondary School Students. Master's thesis, Ouargla University.
- Sareer, Mohammed Sharaf & Khaldi, Nour Al-Din (Eds.). Teaching with Objectives and the Pedagogy of Assessment. Setif: Dar Al-Kitab.
- Shahata, Hassan. (2013). Reference in Teaching and Assessment: Modern Challenges and Creative Insights. Cairo: Dar Al-Alam Al-Arabi.
- Shamsa A. , Munazza M. and Zahra R. (2018). Implementation of CIPP Model for Quality Evaluation at School. Journal of Education and Educational Development Vol. 5 No. 1 PP 189-206
- Taabli, Mohamed Taher & Gouareh, Mohamed. (2013). A Theoretical Approach to the Concept of Achievement Tests and Their Types. Issue 10.
- Tamam, Ismail Tamam, Kamel, Rashdi Fathi, & Amin, Zeinab Mohamed. (1997). Future Trends in Science Teaching and Educational Technology. 1st ed. Assiut: Modern Offset Printing Press.
- Tjabolo, Siti & Otaya, Lian. (2019). The Evaluation of Islamic Education Teachers' Performance. Jurnal Pendidikan Islam. 5. 25-38..
- Yehia, Alwan. (2007). Educational Evaluation and Measurement and Its Role in the Success of the Educational Process. Journal of Humanities Sciences, University of Mohamed Khider Biskra, Issue 11.

- Zaitoun, Ayesh Mahmoud. (2007). Constructivist Theory and Strategies of Science Teaching. 1st ed. Amman: Dar Al-Shorouk for Publishing and Distribution.