

1. alternative form assessment?

Alternative assessments, also referred to as performance tests or authentic assessments, are used to determine what students can and cannot do, in contrast to what they do or do not know. In other words, an alternative assessment measures applied proficiency more than it measures knowledge. Typical examples of alternative assessments include portfolios, project work, and other activities requiring some type of rubric.

What follows are some examples of alternative assessments that you might consider.

- Open book exams. ...
- Crib Sheets. ...
- Take home exams. ...
- Collaborative testing. ...
- Student portfolios. ...
- Performance Tests. ...
- Retake policies. ...
- Adding the option of explanation to an M-C test.

Advantages of Alternative Assessments

- They provide a means of assessing valued skills that cannot be directly assessed with traditional tests.
- They provide a more realistic setting for student performance than traditional tests.
- They focus on student performance and the quality of work performed by students.
- They can be easily aligned with established learning outcomes.
- Process can be costly in terms of time, effort, equipment, materials, facilities, or funds.

2. construct validity?

Meaning

How well a test measures up to its claims. A test designed to measure *depression* must only measure that particular construct, not closely related ideals such as *anxiety* or *stress*.

Procedure

Develop a test framework;

- 1) Defining construct,
- 2) Identifying sub-constructs,
- 3) Listing indicators of each sub-construct
- 4) Writing test items for each indicator

Method

There are two methods to confirm construct validity of a test.

Expert judgment

- 1) There are experts of the field. For above example, people who are expert in essay writing will be considered to assess the construct validity of the table and table will be revised under their guidance.

2) Factor analysis

Criterion-referenced Assessment

Criterion-referenced assessment describes student performance according to a specific domain of clearly defined learning tasks e.g. adds single-digit while numbers. In this you don't compare student performance with other students rather you compare the performance of all students with criteria (in our case that criteria are our learning outcomes). It is most commonly used in schools to report the achievement of learning outcomes against set goals rather than other students. It grades the students to pre-defined criteria and student's grades represent their mastery over content. Students with same level of expertise achieve same level of grades. A cut point is determined to distinguish between failed and successful students regardless of score of highest and lowest achiever. It consists of teacher-made tests, custom made tests from the test publishers and observational techniques.

Function of formative assessment

Functions of Formative Assessment

- 1) Focus of measurement in formative assessment is predefined segment of instruction.
- 2) Limited sample of learning tasks are addressed.
- 3) The difficulty of item varies with each segment of instruction.
- 4) Formative assessment is conducted periodically during the instructional process.
- 5) Results of formative assessment are used to improve and direct learning through ongoing feedback.

3 suggestion for good assessment 3

A good assessment should **be reliable, valid, and free of bias**. First, reliability refers to the consistency of students' scores; that is, an assessment is reliable when it produces stable and consistent results. Reliability can come in two major forms: (1) stability and (2) alternate form reliability.

A quality assessment in education consists of four elements - reliability, standardization, validity and practicality

A good assessment is supposed to show what we have truly learned. There are four qualities of good assessments. Educators should ensure these qualities are met before assessing students. They are:

- Reliability
- Standardization
- Validity
- Practicality

function of blue print test

Table of specification

One of the tools used by teachers to develop a blueprint for the test is called —Table of Specificationll in other words Table of Specification is a technical name for the blue print of test. It is the first formal step to develop a test.

Concept of Table of Specification

- 1) It helps a teacher in allotting the questions to different content areas and Bloom’s learning categories in a systematic manner.
- 2) The blueprint is meant to insure content validity. Content validity is the most important factor in constructing an achievement test. (will be discussed in later unit)
- 3) A unit test or comprehensive exam is based on several lessons and/or chapters in a book supposedly reflecting a balance between content areas and learning levels (objectives).

In developing a test blueprint first of all it is necessary to select some learning. Objectives and among this list of learning objectives some objectives are more important in sense that more time of instruction is spent on them while some other are less important in terms of time spent on them in classroom so in developing table of specification balance among these learning objectives is important, for this purpose we need to weigh the learning objectives for calculating their relative weightage in test.

The test blueprint is used to guide and target item writing as well as for test form assembly. Use of a test blueprint improves consistency across test forms as well as helping ensure that the goals and plans for the test are met in each operational test. An example of a test blueprint is provided next.

In this post, I discuss the benefits of blueprints.

- Create a shared understanding of the bigger picture.
- Understand what different teams are working on.
- Enable better communication and outcomes.
- Help identify and explain priorities.
- Stay focused by zooming in and out.
- Conclusion.

learning student outcomes in national curriculum

Student learning outcome:

- 1) Articulate, identify and differentiate between the sounds of individual letters, digraphs and trigraphs in initial and final positions in a word.
- 2) Identify paragraph as a graphical unit of expansion, $\frac{3}{4}$ know that words in a sentence join to make sense in relation to each other.

Multiple choice questions

The MCQs are the most versatile type of test item available. It can measure variety of learning outcomes from simple to complex, and it is adaptable to the most type of matter content.

Despite the wide applicability of the MCQs there are SLOs, such as ability to organize and present ideas that cannot be measured with any form of selection item. For such skills constructed and restricted response questions are used.

A multiple-choice question (MCQ) is composed of two parts: **a stem that identifies the question or problem**, and a set of alternatives or possible answers that contain a key that is the best answer to the question, and a number of distractors that are plausible but incorrect answers to the question.

Essay type question

Essay question are defined as a test item which require a response composed by the examinee, usually in the form of one or more sentences , of a nature that no single response or pattern of

response can be listed as correct , and the accuracy and quality of which can be judged subjectively only by one skilled or informed in the subject.

writting essay type question steps.

- 1 Step 1: Prewriting. Before you start writing, you need to decide exactly what you'll write about and do the necessary research. ...
- 2 Step 2: Planning and outlining. ...
- 3 Step 3: Writing a first draft. ...
- 4 Step 4: Redrafting and revising. ...
- 5 Step 5: Editing and proofreading.

Advantages of essay type question.

- Test takers can elaborate and provide detailed answers.
- Test takers are not able to guess and select an answer.
- Can review individualized responses from each user.
- Can be used for all types of subjects.
- Takes less time to create questions.

guidelines for essay type questions?(5)

Following guidelines are for writing essay type items when developing a test.

- 1) Frame questions so that the examinee's task is explicitly defined.
- 2) Specify the value and an approximate time limit for each question.
- 3) Do not employ optional questions.
- 4) Employ a larger number of questions that require relatively short answers rather than only a few questions that require long answers.
- 5) Verify a question's quality by writing a trial response to the question.
- 6) Prepare a tentative scoring key in advance of considering examinee responses.
- 7) Score all answers to one question before scoring the next question.
- 8) Make prior decisions regarding treatment of such factors as spelling and punctuation.
- 9) Evaluate essay responses anonymously.

Q: State 5 rules for constructing alternative form questions.

- ANS: 1. Avoid broad general statements if they are to be judged true or false.
2. Avoid trivial statements
 3. Avoid the use of negative especially double negative statements.
 4. Avoid to include two ideas in one statement unless cause and effect relationship are being measured
 5. Avoid long complex sentences

Steps of rubric

Basic Steps to design Rubric

1. Identify a learning goal.
2. Choose outcomes that may be measured.
3. Develop or adapt existing rubric.
4. Share it with students.

2.formula of instructional time of table of specification.

Step 1: Instruction Time

To do the calculations for the instruction time for columns of the table of specifications the teacher must use the following formulas for each objective in the table.

Time in class spent on objective (min) / total time for the instruction being examined (min) Percentage of instruction time

Percentage of instruction time =
$$\frac{\text{time spent on objective, content, theme (min)}}{\text{Total time for the instruction being examined (min)}}$$

Percentage of instruction time =
$$\frac{250}{1000}$$

Percentage of instruction time = 25%

3.what is test-retest. Write two advantage and two disadvantages.

Test-Retest method is a measure of stability

Test-retest reliability measures the consistency of results when you repeat the same test on the same sample at a different point in time. You use it when you are measuring something that you expect to stay constant in your sample.

Arranging test items from easy to hard

- 1. Increases the possibility of good start for poor starters
- 2. Build confidence
- 3. Motivates
- 4. Reduces test anxiety

5- advantage of writing test

Exams help a lot to bring an improvement in the individual’s knowledge because they provide regular feedback to the students who acknowledge their shortcomings and work on them. Another advantage is that exams promote competition among students. They work harder to improve their knowledge and skills.

test security?

Test security is the process of protecting assessments, so that the results from those assessments can be trusted and used to make important decisions about individual competence. According to the American Education Research Association/American Psychological Association/National Council on Measurement in Education standards, test users have the responsibility to protect the security of tests. Test security is particularly important when assessments are used, for example, in selection, accountability, credentialing, or diagnosis because of the inferences drawn from the assessment’s validity.

3)guessing technique for appraisal peer?

Peer appraisal Peer appraisal

In this procedure students rate their peers on the same rating device used by their teacher.

It depends on greatly simplified procedures.

Guess who technique

In this technique teacher use a positive or negative behavior of a student as a example. Other students from same group try to guess the statement with that characteristic correctly. Generally

behaviors used for example are positive in nature to avoid any adverse effect on the student pointed out in the example.

The guess who technique is based on nomination method of obtaining peer ratings and is scored by simply counting the number of mentions each students receive on each description.

4)advantages of MCQs?

Advantages of MCQs

- A. Ensure objectivity, reliability and validity; preparation of questions with colleagues provides constructive criticism.
- B. Increase significantly the range and variety of facts that can be sampled in given time.
- C. Provide precise and unambiguous measurement of the higher intellectual processes.
- D. Provide detailed feedback for both student and teachers.
- E. MCQ are easy and rapid to score.

7)steps to starting general objectives?(5)

Steps for stating General Objectives

- 1) State each general objective as an intended learning outcome
- 2) Begin each general objective with a verb
- 3) State each general objective to include only one general learning outcome
- 4) State each general objective at the proper level of generality
- 5) Keep each general objective sufficiently free of course content so it can be used with various units of study
- 6) Minimize the overlap with other objectives

How parent teacher meeting conduct

5 Steps to Conducting Successful Parent-Teacher Conferences

Parent-teacher conferences are held both for discussing students’ academic progress, and for building and strengthening parent-teacher partnership to achieve academic goals. Here are 5 steps that teachers can follow to execute smooth and productive parent-teacher conferences:

Keep Parents Posted: Ideally, teachers and parents should communicate before the scheduled conference date. Parents must be informed of any issues that need to be discussed – this way, the conference serves as a brainstorming session for finding solutions to improve student performance in class.

Present Evidence: Teachers can create portfolios for each student recording behaviors, class participation, academic progress, achievements, and areas that need development. At conferences, these portfolios can help parents gain a clear and complete picture about their child’s academic progress.

Be Purposeful: Sometimes, the discussion may move away from the student or fixate on a particular point. Teachers can try to avoid this roadblock by forming an agenda for the conference, clearly listing the points that need to be discussed. A copy of the agenda can be shared with parents either before or during the conference and it can help teachers direct parents’ attention back to the necessary points.

When scheduling conference timing, teachers can also plan additional time for miscellaneous topics, or for parents to ask questions or voice concerns. If needed, teachers can schedule an additional meeting to discuss these issues.

Be Positive: Remember to start and end the conference on a positive note. Highlight achievements and progress before discussing the areas that require improvements. Teachers can collaborate with parents and find suitable solutions to help the students with their problems. End the meeting on a positive note, as this can give parents hope and motivate them to help their child.

Define assessment

Assessment is **the systematic basis for making inferences about the learning and development of students**. It is the process of defining, selecting, designing, collecting, analyzing, interpreting, and using information to increase students' learning and development.

Nature of reliability

1. Nature of reliability.

Reliability refers to the consistency of measurement

- 1) Reliability refers to the results obtained with an assessment instrument and not to the instrument itself.
- 2) An estimate of reliability always refers to particular type of consistency (stability, equivalence, internal consistency)
- 3) Reliability is necessary but not sufficient condition for validity.
- 4) Reliability is primarily statistical (range +1 and -1).

Method of assessment

- 1) Placement Assessment.
- 2) Formative Assessment
- 3) Diagnostic Assessment
- 4) Summative Assessment

Benefits of rubric scoring.

- Help clarify vague, fuzzy goals.
- Help students understand your expectations.
- Help students self-improve.
- Inspire better student performance.
- Make scoring easier and faster.
- Make scoring more accurate, unbiased, and consistent.
- Improve feedback to students.
- Reduce arguments with students.

Q: Restricted response type questions with example

ANS: Pose a specific problem for which student needs to recall suitable information, organize it, derive a defensible conclusion, and express it within the given limits of the questions (like 400 word, 30 min. etc.).

Example: List the similarities and differences in the process of cell division in meiosis and mitosis? A variety of learning outcomes (Linn, 2000) can be checked by using this format of essay question. Some of which are;

1. Analysis of relationship.
2. Compare and contrast positions.
3. Explain cause-effect relationship.
4. Organize data and support a viewpoint.
5. Formulate hypotheses.
6. Point out strengths and weaknesses.
7. Integrate data from various resources.

Elements of Rubric

A rubric includes:

- 1. Score
- 2. Criteria
- 3. Level of performance
- 4. Descriptors

Types of written test 2

Types of Written Tests

- ✓ Verbal
- ✓ Non-verbal
- ✓ Objective
- ✓ Subjective
- ✓ Teacher Made
- ✓ Standardized
- ✓ Power
- ✓ Speed

1_Solo stands for? Names of solo taxonomy?

Structure of Observed Learning Outcomes (SOLO)

1.Pre-structural

Students are simply able to acquire bits of unconnected information and respond to a question in meaningless way. Example of pre-structural level:

Question: What is your name?

Answer: What is your name?

Uni Structural

Student shows concrete understanding of the topic. But at this level is only able to respond one relevant element from the stimuli or item that is provided.

Indicative verbs: identify, memorize, do simple procedure

Multi- Structural

Student can understand several components but the understanding of each remains discreet. A number of connections are made but the significance of the whole is not determined. Ideas and concepts around an issue are disorganized and aren't related together.

Indicative verbs: enumerate, classify, describe, list, combine, do algorithms.

Relational

Student can indicate connection between facts and theory, action and purpose. Shows understanding of several components which are integrated conceptually showing understanding of how the parts contribute to the whole. Indicative verbs: compare/contrast, explain causes, integrate, analyze, relate, and apply.

Extended Abstract

Student at this level is able to think hypothetically and can synthesize a material logically. Student make connections not only with in the given subject area but understanding is transferable and generalizable to different areas. Indicative verbs: theorize, generalize, hypothesize, reflect, generate

Note on types of written test 5

Types of Written Tests

- ✓ Verbal
- ✓ Non-verbal
- ✓ Objective
- ✓ Subjective
- ✓ Teacher Made
- ✓ Standardized
- ✓ Power
- ✓ Speed

Verbal

Emphasize reading, writing, or speaking. Most tests in education are verbal tests.

Non-verbal

Does not require reading, writing or speaking ability, tests composed of numerals or drawings is example.

Objective

Refers to scoring of tests when two or more scorers can easily agree on whether the answer is correct or incorrect, the test is objective one. True false, multiple choice and matching tests are example

Subjective

Also refers to scoring. When it is difficult for two scorers to agree on whether an item is correct or incorrect, the test is a subjective one. Essay tests are the example.

Teacher Made

Constructed solely by teacher only to be used in his/her own classroom. This type of test is custom designed according to need and issues related to specific class

Standardized

Test constructed by measurement experts over a period of years. They are designed to measure broad national objectives and have a uniform set of instructions that are adhered to during each administration

Most also have tables of norms, to which a student performance may be compared to determine where the student stands in relation to a national sample of students at same level of age or grade

Power

Tests with liberal time limits that allow each student to attempt each item. Item tend to be difficult.

Speed

Tests with time limits so strict that no one is expected to complete all items. Items tend to be easy.

Difference between summative and formative assessment

Formative Assessment

Formative assessment provides feedback and information during the instructional process, while learning is taking place, and while learning is occurring. Formative assessment measures student progress but it can also assess your own progress as an instructor.

Types of Formative Assessment

- 1) Observations during in-class activities; of student's non-verbal feedback during lecture.

- 2) Homework exercises as review for exams and class discussions)
- 3) Reflections journals that are reviewed periodically during the semester
- 4) Question and answer sessions, both formal (planned) and informal (spontaneous)
- 5) Conferences between the instructor and student at various points in the semester
- 6) In-class activities where students informally present their results
- 7) Student feedback collected by periodically answering specific question about the instruction and their self-evaluation of performance and progress

Functions of Formative Assessment

- 6) Focus of measurement in formative assessment is predefined segment of instruction.
- 7) Limited sample of learning tasks are addressed.
- 8) The difficulty of item varies with each segment of instruction.
- 9) Formative assessment is conducted periodically during the instructional process.
- 10) Results of formative assessment are used to improve and direct learning through ongoing feedback.

Summative Assessment

Summative assessment takes place after the learning has been completed and provides information and feedback that sums up the teaching and learning process. Typically, no more formal learning is taking place at this stage, other than incidental learning which might take place through the completion of projects and assignments.

Summative assessment is more product-oriented and assesses the final product, whereas formative assessment focuses on the process toward completing the product. Once the project is completed, no further revisions can be made.

If, students are allowed to make revisions, the assessment becomes formative.

Types of Summative Assessment

- ✓ Examinations (major, high-stakes exams)
- ✓ Final examination (a truly summative assessment)
- ✓ Term papers (drafts submitted during the semester would be a formative assessment)
- ✓ Projects (project phases submitted at various completion points could be formatively assessed)
- ✓ Portfolios (could also be assessed during its development as a formative assessment)
- ✓ Performances

Functions of Summative Assessment

- 1) Focus of measurement in summative assessment is on course or unit objectives.
- 2) Broad sample of all objectives is used in summative assessment.
- 3) This type of assessment uses wide range of difficulty while selecting items for the test.
- 4) Summative assessment is done at the end of the unit or the course
- 5) Most important functions of summative assessment is to assign grade, certification of accomplishment and evaluation of teaching

5.Suggestions for scoring constructed response questions.5 marks

Suggestions for scoring constructed response questions

1. Prepare an outline of the expected answer in advance.
2. Use the scoring rubric that is most appropriate.
3. Decide how to handle factors that are irrelevant to the learning outcomes being measured.

- Evaluate all responses to one question before going on to the next one.
- When possible, evaluate the answers without looking at the student's name.
- If especially important decisions are to be based on the results, obtain two or more independent ratings.

6.steps for stating general objectives. 5 marks.

Steps for stating General Objectives

- State each general objective as an intended learning outcome
- Begin each general objective with a verb
- State each general objective to include only one general learning outcome
- State each general objective at the proper level of generality
- Keep each general objective sufficiently free of course content so it can be used with various units of study
- Minimize the overlap with other objectives

Derived score 2

Derived Scores

A derived score is a numerical report of test performance on a score scale that has welldefined characteristics and yields normative meaning.

Holistic rubric 3

Holistic Scoring Rubric

These rubrics are good for evaluating overall performance on a task. All criteria are assessed as a single score. As only one score is given so holistic rubrics are easy to score.

When to use:

- There is no correct answer/response to ask a task e.g. creative work.
- Focus is on overall quality, proficiency or understanding of a specific content or skill.
- The assessment is summative e.g. at the end of the semester or major.
- Assessing significant numbers e.g. 150 student portfolio.

5 recommendations for effective HST 5

Recommendations for effective High Stake Testing

- Protection against high stake decisions based on single test
- Adequate resources and opportunity to learn
- Validation for each intended separate use
- Full disclosure of likely consequences
- Alignment between test and curriculum
- Validity of passing scores and achievement levels
- Appropriate attention towards language difference between examinees
- Appropriate attention towards examinees with disabilities
- Careful adherence to explicit rules for determining which students are to be tested
- Ongoing evaluation for intended and unintended effects of high stake testing

State the steps of specific learning objectives 5

Specific Learning Outcomes

Each General objective must be defined by a sample of specific learning outcome to clarify how students can demonstrate that they have achieved general objective. Until the general objective are further defined in this manner they will not provide adequate direction for assessment

Steps for Stating Specific Outcomes

- 1) List beneath each general objective a representative sample of specific learning outcome that describes terminal performance students are expected to demonstrate
- 2) Begin each specific learning outcome with an action verb that specifies observable performance
- 3) Make sure that each specific learning outcome is relevant to the general objective it describes
- 4) Include enough SLOs to describe adequately the performances of students who have attained the objectives.
- 5) Keep the SLOs sufficiently free of course content so that the list can be used with various units of study
- 6) Consult reference materials for the specific components of those complex outcomes that are difficult to define
- 7) Add a third level of specificity to the list of outcomes, if needed

What is Mcqs?2

Multiple Choice Items

The multiple choice items is generally recognized as the most widely applicable and useful type of objectives test items that can measure knowledge, comprehension and as well as application level learning outcomes.

Multiple choice items consist of two parts:

A Stem

No. of options or alternatives

Stem

The problem may be stated as direct question or an incomplete statement and is called the stem of the item.

SOLO stands for and write its levels? 2

Structure of Observed Learning Outcomes (SOLO)

Levels of Structure of Observed Learning Outcomes (SOLO)

- 1) Pre-structural
- 2) Uni-structural
- 3) Multi-structural
- 4) Relational
- 5) Extended Abstract

Norm-referenced Assessment 5 marks

Norm-referenced Assessment measures the students' performance according to relative position in some known group. E.g. ranks tenth in classroom group of 50 or top 5 students in a class. Relative

position means a point defined with reference to another position (where a student stands compared to other students)

Rather than reporting student’s achievement it reports students standing among other students.

NRT is utilized to discriminate between a certain groups of students. It’s never used for certification or issuing grades to students. Position of the student is generally represented by percentile scarce noting the percentage of students achieving same or low scarce in the test. Examples of NRT are NTS or CSS exams. E.g. if you achieved 97% percentile on NTS test, it means there are 96% who scored lower then you.

Norm referenced test includes items with average difficulty and high discriminating power. This provides large spread of scores which makes it easy to declare the relative positions of students. Because the purpose of NRT is not to certify grades, so the test made for it must be of average difficulty means test items should not be very easy or very tough. If all the items are difficult then no student will be able to solve it and we are not able to discriminate who is good students. And if all the items are easy then even the low ability students can solve it then we are not able to discriminate it.

Norm referenced test includes standardized aptitude and achievement tests, teacher-made survey tests, interest inventories, adjustment inventories

Table of specification.

One of the tools used by teachers to develop a blueprint for the test is called —Table of Specificationll in other words Table of Specification is a technical name for the blue print of test. It is the first formal step to develop a test.

Nature of validity-----5

Following are different points that describe the nature of validity.

- 1. Appropriateness of the interpretation of the results

Validity is referred as validity of testll but it is in fact validity of the interpretation and use to be made of the results.

- 2. Validity is matter of degree

It does not exist on all or none basis. It is best considered in term of categories that specify degree, such as high, moderate or low validity

- 3. Validity is specific to some particular use or interpretation

No assessment is valid for all purposes. An arithmetic test may have high degree of validity for computational skill and low degree for arithmetical reasoning

- 4. Validity is a unitary concept

Validity does not have different types. It is viewed as a unitary concept based on different kind of evidences

- 5. Validity involves an overall evaluative judgment

Portfolio

Systematic collection of students work into portfolios can serve a variety of instructional and assessment purposes. The value of portfolios depend heavily on the clarity of purpose the guidelines for the inclusion of materials, and the criteria to be used in evaluating portfolio.

Portfolio of students work

A portfolio is collection of student work selected to serve a particular purpose such as documentation of student growth. It is purposeful; collection of pieces of student's work.

Key steps in defining and using portfolios

- 1. Specify purpose.
- 2. Provide guidelines for selecting portfolios.
- 3. Define student's role in selection and self-evaluation.
- 4. Specify evaluation criteria.
- 5. Use portfolios in instruction and communication.

Strengths of portfolios

- 1) The can be readily integrated with the instruction.
- 2) Provide opportunity to student's to show what they can do.
- 3) Encourage to become reflective learner.
- 4) Help in setting goal and self-evaluation
- 5) Help teacher and student to collaborate and reflect on student's progress.
- 6) Effective way to communicate with parents.
- 7) Provide mechanism for student centered and student-directed conferences with parents.
- 8) Provide concrete examples of students development and current skills.

Weaknesses of portfolios

- 1) Can be time consuming to assemble.
- 2) Hard to use in summative assessment.
- 3) Difficult to compare results.
- 4) Very low reliability.

Evaluation 2 marks

Evaluation is process of making a value judgment against intended learning outcomes and behavior, to decide quality and extent of learning. Evaluation is always related to your purpose, you aligned your purpose of teaching with what students achieved at the end, with their quality and quantity of learning

Classical Test Theory (CTT) and Item Analysis 5

Classical Test Theory (CTT) has relatively weak theoretical assumptions, which make it easy to apply in many testing situations. Relatively weak theoretical assumptions not only characterize CTT but also its extensions (e.g., generalizability theory).

CTT's major focus is on test-level information, item statistics (i.e., item difficulty and item discrimination) are also an important part of the CTT model.

Guidelines for essay type items

- Following guidelines are for writing essay type items when developing a test.

- Frame questions so that the examinee's task is explicitly defined.
- Specify the value and an approximate time limit for each question.
- Do not employ optional questions.
- Employ a larger number of questions that require relatively short answers rather than only a few questions that require long answers.
- Verify a question's quality by writing a trial response to the question.
- Prepare a tentative scoring key in advance of considering examinee responses.
- Score all answers to one question before scoring the next question.
- Make prior decisions regarding treatment of such factors as spelling and punctuation.
- Evaluate essay responses anonymously

Reliability 3 marks

Reliability

Reliability refers to the consistency of assessment results.

Reliability refers to the results obtained with an assessment instrument and not to the instrument itself.

An estimate of reliability always refers to particular type of consistency (stability, equivalence, internal consistency)

Reliability is necessary but not sufficient condition for validity.

Reliability is primarily statistical (range +1 and -1).

Methods of reliability

Test-Retest (stability)

Equivalent Forms (equivalence)

Test- Retest with Equivalent Forms (stability and equivalence)

Split Half (Internal Consistency)

Kuder- Richardson(Internal Consistency)

Cronbach Alpha (Internal Consistency)

Inter-rater Reliability (consistency of rating)

Validity

Validity is an evaluation of adequacy and appropriateness of the interpretation and uses of results. It determines if a test is measuring what it intended to measure.

Reliability vs Validity

Reliability of measurement is needed to obtain the valid results, but we can have reliability without validity. Reliability is necessity but not sufficient condition for validity

Q: analytic rubric? In which conditions it is used?

ANS: Analytic rubric

Each criterion is assessed separately by using different descriptive ratings. Each criterion is given a separate score. Final score is made up of adding each component parts. It takes more time to score but gives detailed more feedback.

When to use:

- 1. Several faculty are collectively assessing student work.
- 2. Outside audience will be examining rubric scores.
- 3. Profiles of specific strength/weakness are desired.

Q: Extended response type questions with example

ANS: It allows the students to determine the length and complexity of the response. This type of questions is most suitable for measuring higher level mental processes skills like synthesis and evaluation. Teachers prefer these questions when they want to have information about student's ability to organize, integrate, express and evaluate information.

Example: What are the teachings of Islam about the respect of elders? How do these teachings help us to make a good society? Marks will be given for correct information, your own point of view, good judgment and good presentation.

Q: assembling or packaging the test

ANS: Assembling Test Once you have measureable instructional objectives, test blueprint, written test items matching instructional objectives, so you are ready to

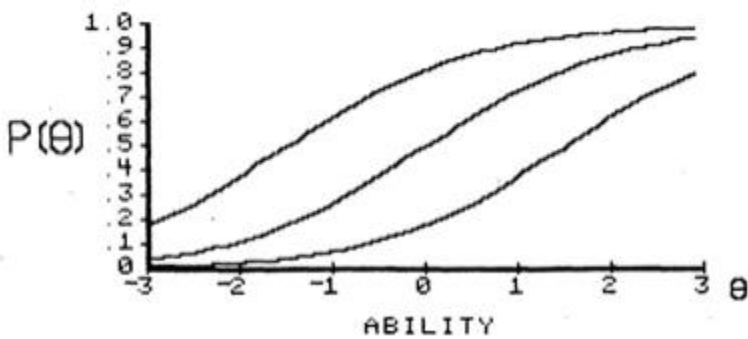
- package the test and • reproduce the test

Assembling Test (Packaging the Test)

- 1. Grouping together items of similar format
- 2. Arranging test items from easy to hard
- 3. Properly spacing items
- 4. Keeping items and options on the same page
- 5. Placing illustrations near the descriptions
- 6. Checking the answer key

Q: What is Item difficulty in IRT? Explain with graph.

ANS: The application of Item difficulty in IRT is defined as the ability at which the probability of success on the item is .5 on a logit scale, which is also known as threshold difficulty (see Fig. 3). An item's level of difficulty is another factor affecting an individual's probability of responding in a particular way. An item that has a high level of difficulty will be less likely to be answered correctly by an examinee with low ability than an item that has a low level of difficulty (i.e., an easy item).



Score

A system of number or values assigned to a work often combines with a level of performance. High numbers are for best performance like 4, 5 or 6 whereas down to 1 or 0 are the lowest score in a performance assessment.

Criteria

It tells us that which feature, trait or dimension to be measured and include a definition and example to make clear the meaning of each trait to be assessed.

Level of Performance

There are adjectives to describe the performance levels. These levels tell students what they are expected to do.

Descriptors can be used with levels of performance to achieve objectivity but they can be used without them as well.

Descriptors

These are details for each level of performance to reliable and unbiased scoring.

Formative Assessment

Formative Assessment determines learning progress, provides feedback to reinforce learning, and correct learning errors. When we assess student during classroom instruction with a purpose to have a feedback that how can we make our teacher learning process better, that is formative assessment. In this assessment, we are not assessing what students learnt or not rather we assess the process behind the students learning .The process behind the student learning includes teaching method , book, . If we make all these things according to the needs of students then learning will improve.

Summative Assessment-----5

Summative assessment comes at the end of the instructional session (course of unit). It is designed to measure the extend of achievement of intended learning outcomes. The primary utility of this type of assessment is to assign grades and certifying the level of mastery and expertise in certain subject. It is usually done through teacher made achievement tests or alternative assessment techniques like portfolio to summarize the overall performance of the student at the end of the session. It is not compulsory to done it at the end of the semester, in semester system there is midterm and final term in one semester, and these both are summative assessment.

Sociometric technique

Sociometric technique is method for assessing the social acceptance of individual students and the social structure group. It is based on students choice of companion for some group situation.

This form was used to measure student’s acceptance as seating companions, work companions and play companions.

There are few important principles of sociometric choosing

71

- 1. The choices should be real choices that are the natural part of classroom activities.
- 2. The basis for the choice and restriction on the choosing should be made clear.
- 3. All students should be equally free to participate in the activity or situation.

- 4. The choice of each student make must be kept confidential.
- 5. The choices should actually be used to organize or rearrange the group

Name of Psychomotor Domain

- i. Perception ii. Set
- iii. Guided Response iv. Mechanism
- v. Complex covert Response vi. Adaption
- vii. Origination

Theory (IRT) and Item Analysis

Item Response Theory (IRT) presents a model for expressing the association between an individual's response to an item and the underlying latent variable (often called "ability" or "trait") The latent variable, expressed as theta (θ), is a continuous one-dimensional construct that explains the covariance among item responses People at higher levels of latent trait have a higher probability of responding correctly or endorsing an item.

High stake testing

- The use of test and assessment to make decisions that are of prominent educational, financial, or social impact.
- Decision based on high stake testing
- Promotion to next grade
- Awarding of diploma or degree
- Evaluation of school performance
- Incentives and accountability of school staff

Holistic

Holistic rubric is probably more appropriate for extended response essay type items involving a student's abilities to synthesize and create and when no definite answer can be prespecified.

Bloom taxonomy

Bloom's Taxonomy of Learning Objectives Bloom's Taxonomy was presented by Benjamin Bloom in the 1956, consists of a framework with most common objectives of classroom instruction.

Bloom's Taxonomy of Learning Objectives

Those dealing in three different domains and further sub categories in these domains.

Cognitive

Affective

Psychomotor

Cognitive Domain:

Knowledge

Comprehension

Application

Analysis

Synthesis

Evaluation

Affective Domain

Receiving

Responding

Valuing

Organization

Characterization

Psychomotor Domain

Perception

Set

Guided Response

Mechanism

Complex covert Response

Adaption

Origination

High Stake Testing-----5

Why does it matter?

The use of test and assessment to make decisions that are of prominent educational, financial, or social impact.

Decision based on high stake testing

- 1. Promotion to next grade

- 2. Awarding of diploma or degree
- 3. Evaluation of school performance
- 4. Incentives and accountability of school staff

High Stake Testing in Pakistan

In school education in Pakistan, high stake testing is conducted at each level from primary to higher secondary.

- 1. Grade 5 and 8 examination by Punjab Examination Commission.
- 2. Grade 10 and 12 examination by BISEs

Test construction in High Stake testing

- 1. The format of items are same as used in CRT in classroom
- 2. Item of developed by professional employed in dedicated organization

Item banks are developed

- 4. Psychometric properties of item are tested
- 5. The process goes on round the year

Criticism on high stake testing

High stake tests have same issues as classroom test but much larger impact and consequences.

Example of knowledge comprehension and application level

Knowledge:

It is defined as the remembering of previously learned material. This may involve the recall of a wide range of facts, procedures principals and generals, the recall of procedures and the processes.

Sample Question: Define the 6 levels of Bloom's taxonomy of the cognitive domain.

Comprehension:

It is defined as the ability to grasp the meaning of the material. Individual can make use of the content or idea being communicated without necessarily related it to other content and seeing its fullest implications.

Sample Question: explain the purpose of Bloom's taxonomy of the cognitive domain.

Application:

It refers to the ability to use the previously learned material in new and concrete situations. The abstractions may be in the shape of universal ideas, rules of methods.

Sample Question: write an instructional objective for each level of Bloom's taxonomy.

Item characteristic curve

To analyze items using IRT, the main thing need to consider is item characteristic curve (ICC). The item characteristic curve is considered as the basic building block of item response theory.

- It gives a picture of:
- The item difficulty
- Discrimination power
- The probability of answering correctly by guessing

Advantages of anecdotal record:

- Following are the advantages of anecdotal records.
- It depicts actual behaviors in natural situations.
- Facilitate gathering evidence on the events that are exceptional but significant.
- Beneficial for students with less communication skills.

Steps to design rubric

- Basic Steps to design Rubric
- Identify a learning goal.
- Choose outcomes that may be measured.
- Develop or adapt existing rubric.
- Share it with students

Test security

Educational decisions

In the classroom, decisions are constantly being made. Teachers face huge numbers of dilemmas every day. These decisions can be of following nature

Types of Educational Decisions

- Instructional
- Grading
- Diagnostic
- Selecting
- Placement
- Counseling and Guidance
- Program or Curriculum
- Administrative

These types of decisions are taken at different levels. Some are decided at Board/Administrative level while some are taken at school management level and other are taken in classrooms by teachers

Scoring Rubrics for Essay Type Items

Scoring of essay items is a time consuming and difficult process. Reliability of the test demands that scoring should be consistent not only by the rater at different times but by two independent raters as well.

Scoring rubrics are descriptive scoring schemes that are developed by teachers or other evaluators to guide the analysis of writing products or processes of student's piece of writing or any performance

Alternatives

The list of suggested solutions which may include words, numbers, symbols or phrases are called alternatives (also called choices or options).

The correct alternative in each item is called the answer while all incorrect or less appropriate alternatives are called distractors or foiled and the student's task is to select the correct or best alternative from all the options.

Formative assessment

Formative assessment provides feedback and information during the instructional process, while learning is taking place, and while learning is occurring. Formative assessment measures student progress but it can also assess your own progress as an instructor.

Types of formative assessment

- Observations during in-class activities; of student's non-verbal feedback during lecture.
- Homework exercises as review for exams and class discussions)
- Reflections journals that are reviewed periodically during the semester
- Question and answer sessions, both formal (planned) and informal (spontaneous)
- Conferences between the instructor and student at various points in the semester
- In-class activities where students informally present their results
- Student feedback collected by periodically answering specific question about the instruction and their self-evaluation of performance and progress

Written test

In classroom assessment different forms of assessments are utilized. Each form of test has its own benefits and disadvantages. Most common type of assessment used in classrooms is written assessment

Analytical rubric

Each criterion is assessed separately by using different descriptive ratings. Each criterion is given a separate score.

Final score is made up of adding each component parts. It takes more time to score but gives detailed more feedback.

Analytic scoring requires a list of major elements that students are expected to include in an ideal answer of a question.

Analytic rubric is used to score an essay type of question on different points and significantly it works best on the restricted response type essay question.

Advantages

- More detailed feedback.
- Scoring more consistent across students and grades.

Disadvantages

- Time consuming to score.

Norm referenced and criterion reference

NRT

Type of test which tells us where a student stands compared to other students. It helps determining a student's place or rank among a group of similar students. Such kind of test is called norm-referenced test

Dimensions

- It provides estimate of ability in a variety of skills in much shorter time.
- NRT tend to be general. It measures variety of skills at same time but fails to measure them thoroughly.
- It's hard to make decisions regarding the mastery of student's skill in subject

Criterion Referenced Assessment

A second type of test tells us about student's level of proficiency in or mastery of some skill or set of skills. This is achieved by comparing a student's performance to a standard mastery called a criterion. Test that yields such information is called Criterion Referenced Test

Dimensions

- CRT tends to be specific. It measures particular set of skill at one time and focus on level of achievement of that skill. CRT gives clear picture regarding the mastery of student's skill in subject.
- It measures skill more thoroughly so naturally it takes more time comparing to NRT in measuring the mastery of said skill
- Items included in CRT are relatively easier. Around 80% of the students are expected to respond item correctly in the test
- CRT compares students 'performance to standards indicative of mastery.
- Breadth of content sampled is narrow and covers very few objective

Q: what are subjective type questions? Tell names

ANS: Subjective test items (essay questions) are constructed response type questions that can be the best way to measure the students' higher order thinking skills, such as applying, organizing, synthesizing, integrating, evaluating, or projecting while at the same time providing a measure of writing skills. Essay items can vary from very lengthy (5-10 pages), open-ended to limited or restricted response (one page or less). Therefore, essay type questions are divided into two types

- I. Restricted Response Items
- II. Extended Response Items

Q: High stake testing matters

ANS: High-stakes testing as the use of a summative test or an assessment (i.e., a test designed to measure student achievement after a period of instruction has been completed) to make decisions that are of prominent educational, financial, or social impact.

Types of High stake test

Criterion Referenced Test

Norm-Referenced High-Stake Tests

Q: How to design rubric?

- ANS:**
1. Identify a learning goal.
 2. Choose outcomes that may be measured.
 3. Develop or adapt an existing rubric.
 4. Share it with students

Q: Restricted response items names

ANS: Restricted response item: Pose a specific problem for which student needs to recall suitable information, organize it, derive a defensible conclusion, and express it within the given limits the questions (like 400 word, 30 min. etc.)

Some of which are;

1. Analysis of relationship.
2. Compare and contrast positions.
3. Explain cause-effect relationship.
4. Organize data and support a viewpoint.
5. Formulate hypotheses.
6. Point out strengths and weaknesses.
7. Integrate data from various resources

3 marks

Q: Advantage if essay type questions

Ans: Extended response questions have following advantages over other types of question formats:

1. Effective for assessing higher order abilities: analyses synthesize and evaluate.
2. It is comparatively less time consuming to develop such items.
3. Emphasizes essential communicational skills.
4. Guessing is eliminated

Q: State any three limitations of mcqs

Ans: Limitations of MCQs.

1. Take a long time to construct in order to avoid arbitrary and ambiguous questions.
2. Also require careful preparation to avoid preponderance of questions testimony recall.
3. Provide cues that do not exist in practice.

Q: Benefits of rubric in evaluation process

ANS: A rubric is an explicit set of criteria used for assessing a particular type of work or performance and provides more details than a single grade or mark. It is a set of scoring guidelines for evaluating student work.

Q: Item analysis

ANS: The item analysis is an important phase in the development of a test. In this phase statistical methods are used to identify any test items that are not working well. If an item is too easy, too difficult, failing to show a difference between skilled and unskilled examinees, or even scored incorrectly, an item analysis will reveal it.

5 marks:

Q: How self assessment should be implemented

ANS: Boud (1995) argues that the way in which self-assessment is implemented is critical to its acceptance by students. According to Boud (1995), the implementation process needs to include:

- A clear rationale: what are the purposes of this particular activity?
- Explicit procedures—students need to know what is expected of them.
- Reassurance of a safe environment in which they can be honest about their own performance without the fear that they will expose information which can be used against them.
- Confidence that other students will do likewise, and that cheating or collusion will be detected and discouraged (Boud, 1995, p.182).

Q: What is report? Name of reports?

ANS: Reporting is the process of communicating the knowledge and understanding gained from assessing a student’s work to improve learning

There are two forms of reports 1. Formal reports 2. Informal reports

Q: Explain formal and informal report

ANS: Formal reports

Formal reports communicate about important aspects of the students’ progress in the areas of academic, social, human and career development to parents and students.

Informal reports

Informal report tells parents about what their child is able to do. It provides information about the areas of learning that need further attention or development and the ways the teacher is supporting the student’s learning needs.

Q:

ANS: A derived score is a numerical report of test performance on a score scale that has well-defined characteristics and yields normative meaning. Most common type of derived scores are 1. Grade norms

2. Percentile norms

3. Standard norms

ANS: Subjective test items (essay questions) are constructed response type questions that can be the best way to measure the students' higher order thinking skills, such as applying, organizing, synthesizing, integrating, evaluating, or projecting while at the same time providing a measure of writing skills. Essay items can vary from very lengthy (5-10 pages), open-ended to limited or restricted response (one page or less). Therefore, essay type questions are divided into two types I. Restricted Response Items

II. Extended Response Items

ANS: Q: List the similarities and differences in the process of cell division in meiosis and mitosis?

ANS: Q1. Give examples to identify different forms of governments, their advantages and disadvantages over each other. Which form of government is most suitable in the socioeconomic and cultural context of our country? Your response will be graded on accuracy of information, power of arguments, conciseness and orderliness of presentation. It is suggested to keep your response limited to 5000-6000 words.

Q2. What are the teachings of Islam about the respect of elders? How do these teachings help us to make a good society?

ANS: Basic Steps to Design Rubric

1. Identify a learning goal.
2. Choose outcomes that may be measured.
3. Develop or adapt an existing rubric.
4. Share it with students.

- ANS:**
1. Make plans for the meeting.
 2. Begin the meeting in a positive manner.
 3. Present the student’s strong points before describing the areas needing improvement.
 4. Encourage parents to participate and share information.
 5. Plan a course of action cooperatively.
 6. End the meeting with a positive comment.
 7. Use good human relation skills during the meeting

ANS: By format of assessment

- i. Fixed Choice Assessment
- ii. Complex Performance Assessment

Fixed Choice Assessment

Fixed Choice Assessment is used to measure the skills of people efficiently (means measure more skills in less time) and for this we usually use fixed choice items i.e. Multiple choice question, matching exercise, fill in the blanks and true false. It is called fixed choice because the person who is attempting the paper does not need to write the answer, just need to choice the answer.

Complex Performance Assessment

Complex Performance assessment is used for measurement of performance in contexts and the problems valued in their own right. This includes hands on laboratory experiments, projects, essays, oral presentations.