

Objectives

1. **SSC stands for?** (Secondary School Certificate)
2. **HSSC stands for** (Higher Secondary School Certificate)
3. **ID (Item Description)**
4. **IRT** (item response theory)
5. **CRT** (classical response theory)
6. **CTT** (classical test theory)
7. **HST** (High Stake Testing)
8. **IBCC** (Inter board committee of chairman)
9. **BISEs** (Board of intermediate and secondary education)
10. **NEAS** (National Education Assessment System)
11. **PEAS** (Provincial Education Assessment Centre)
12. **EC** (Education Commission)
13. **PEC** (Punjab Education Commission)
14. **DFID** (Development for International Development)
15. **Widely perceived psychosocial theories.** 2
16. **Standardized achievement tests are frequently used in survey test batteries.**
17. **Tests conducted at each level from primary to higher secondary. (High stake testing)**
18. **In case of ____, interpretations can only be made with reference to the constructions on which test was based by the developer. (Standardized test)**
19. **For which test we should take advantage of whatever the material is available. (High stake testing)**
20. **A raw score is a ____ summary of a student's test performance. (numerical)**
21. **In true or false statements one should avoid (Trivial) Statement**
22. **Matching** exercise is used when measuring a student's ability to identify the relationship between a set of similar items, each of which has two components.
23. **Matching is restricted to the measurement of factual information, based on rote learning.**
24. **Short answers are presented in the form of direct questions while completion items are incomplete statements.**
25. **Short answers and completion and fill in the blanks, under the range of objective test items.**
26. **In short answer test item teachers can control possibility of the guessing.**
27. **(Classical test) theories in Psychosocial measurement is known as true score theory.**
28. **In High Stake Testing take advantage of whatever preparation material is available.**

29. In school education in Pakistan, **high stake** testing is conducted at each level from primary to higher secondary.
30. A derived score is a **numerical** report of test performance on a score scale that has well-defined characteristics and yields normative meaning
31. In case of **standardized** test, interpretation can only be made with reference to the constructions on which test was based by the developer.
32. Standardized achievement tests are **frequently** used in the form of survey test batteries.
33. Standardized Achievement Tests are **highly** technical.
34. Standardized Achievement Tests typically are **norm** referenced tests.
35. A system of number or values assigned to a work often combines with a level of performance. **(Score)**
36. Holistic rubric is probably more appropriate for **essay type** items
37. A rubric is an **explicit** set of criteria used for assessing a particular type of work or performance
38. The preparation of test items is greatly facilitated if the items are **properly** recorded.
39. **Guessing** is a standard test-taking strategy presented to examinees taking a multiple choice assessment.

(Subjective)

Q. Types of essay question

Essay type questions are divided into two types

1. Restricted Response Items

- ✚ Statements should not be quoted directly from the text.
- ✚ Evaluate essay responses anonymously.
- ✚ Frame questions so that the examinee's task is explicitly defined.

2. Extended Response Items

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 process skills like synthesis and evaluation.

Q. 3 types of exam body in pk

1. **Examination commission** conduct examination at grade 5 and 8 level
2. **BISEs** hold SSC and HSSC annual examinations

3. **Boards of Technical Education** conduct examination of various diplomas and certificates etc.

Q. What is difficulty in IRT

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. 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).

Q. 3 limitations of true false

Most commonly observed limitations are:

- i. All learning outcomes cannot be measured through alternative form questions. They are generally limited to lower level learning outcomes.
- ii. Ease of guessing correct answers when the answer is not known. With only two choices\ (true or false) the student could expect to guess correctly on half of the items for which correct answers are not known.
- iii. There is sometimes a tendency to take quotations from the text with a minor change in wording.
- iv. There may also be a tendency to include trivial material from the text.
- v. True-false items are prone to high guessing and can only be used for measuring lower level learning outcomes

Q. 3 limitations of matching exercises

- It is restricted to the measurement of factual information, based on rote learning.
- It is highly susceptible to the presence of irrelevant clues.
- Difficulty of findings homogenous material that is significant from the viewpoint of our objectives and learning outcomes

Q. limitations of essay test items

1. The scoring is unreliable and time consuming.
2. Limited sampling of the content is possible.

Q. Define achievement test

A standardized achievement test a fixed set if items to measure defined achievement domain, specific direction for administrating and scoring test, and norms based on representative groups of individuals.

Most published achievement tests are called standardized achievement tests. These typically are norm-referenced tests. Quite a few criterion-referenced achievement tests are also published. Achievement tests are used as part of a broader assessment system or alone. They provide a relatively inexpensive means of measuring broad achievement goals. Standardized achievement tests are often customized to include characteristics of both norm and criterion-referenced tests. Standard content and procedure makes it possible to give an identical test to individuals in different places at different times. Equivalent forms are included in many standardized tests, which make it possible to repeat the test without fear that the test takers will remember the answers from first testing.

Q. 3 suggestions for good short answer question

Word the item so that the required answer is both brief and specific. Do not take statements directly from textbooks to use as a basis for short answer items. A direct question is generally more desirable than an incomplete statement. If the answer is to be expressed in numerical units, indicate the type of answer. When completion items are used, do not include too many blanks or start with a blank space. Also avoid asking trivial information in blank space.

Q. Define NEAS

The National Education Assessment System has been institutionalized in Pakistan at national level with the cooperation of provincial and area Assessment Centers. NEAS was established as a five-year development project with the financial assistance of the World Bank and Development for International Development (DfID) in year 2003. NEAS is a subordinate office under the ministry of Federal Education & Professional Training.

Q. Issues in test administration

1. Cheating
2. Poor testing conditions
3. Test anxiety
4. Errors in test scoring procedure.

Q. Guideline for writing essay type

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

- i. Frame questions so that the examinee's task is explicitly defined.
- ii. Specify the value and an approximate time limit for each question.
- iii. Do not employ optional questions.
- iv. Employ a larger number of questions that require relatively shorter answers rather than only a few questions that require long answers.
- v. Verify a question's quality by writing a trial response to the question.
- vi. Prepare a tentative scoring key in advance of considering examinee responses.

- vii. Score all answers to one question before scoring the next question.
- viii. Make prior decisions regarding treatment of such factors as spelling and punctuation.

ix. Evaluate essay responses anonymously.

Q. Monitor students test invigilation?

- 1) Penalties of cheating
- 2) Disturbing others
- 3) Seating position/posture
- 4) Protecting test from others
- 5) Cheating material handling
- 6) Jurisdiction of invigilation staff

Q. Limitations of short question

- A potential problem with both short-answer and completion items is that difficult to frame questions for one specific answer unless the items are well written.
- Not usable to measure complex learning outcomes.

Q. Uses of matching exercise

The typical matching exercise is limited to measuring factual information, based on simple associations. It is a compact and efficient method of measuring such simple knowledge outcomes. It is also being used with pictorial materials in relating pictures and words to identify positions on maps, charts and diagrams. Regardless of the form of presentation, the student's task is essentially to relate two things that have logical association. This restricts the use of matching exercise to small area of student's achievement.

Q. Score definition.

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.

Q. 3 strategies to solve issue in test administration .

It is equally important to control all the factors other than test itself, to collect trustable evidence of student learning by addressing test administration issues. The way in which the test is administered is very important to meet the goal of producing highly valid, reliable results. Once a test is ready then next step is to administer it. Teacher has to help the students psychologically by maintaining a positive test-taking attitude, clarifying the rules; penalties for cheating, reminding them to check their copies, minimizing distractions and giving time warnings. Cheating, poor testing conditions, and test anxiety, as well as errors in

test scoring procedures contribute to invalid test results. Accurate achievement data are very important for planning curriculum and instruction.

Test scores that overestimate or underestimate students' actual knowledge and skills cannot serve these important purposes. So it is worth a little more time to properly assemble and administer a test.

Q. Scoring rubric definition.

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. Scores levels identified in a scoring rubric must be descriptive, not merely judgmental in nature. Example: Define the level of rubric as —Writing is clear and thoughts are complete as compared to —excellent.

Q. Suggestions for true false

Most important task in formulating statements is free from ambiguity and irrelevant clues. There are list of things to avoid when phrasing the statement

- Avoid trivial statements
- Avoid the use of negative especially double negative statements.
- Avoid long and complex sentences.
- Avoid including two ideas in one statement unless cause and effect relationship are being measured.

Q. Limitation of matching columns?

- It is restricted to the measurement of factual information, based on rote learning.
- It is highly susceptible to the presence of irrelevant clues.
- Difficulty of findings homogenous material that is significant from the viewpoint of our objectives and learning outcomes.

Q. Two uses of true and false?

For measuring such relatively simple learning outcomes, a single declarative statement is used with any one of several methods of responding. One of the most useful functions of true and false items is in measuring the students' ability to distinguish fact from opinion. Another aspect of understanding that can be measured by true and false item is ability to recognize cause and effect relationship. This type of item usually contains true propositions in one statement, and the student is to judge whether the relationship between them is true or false.

Q. short note on matching column?

The matching type exercises consist of two parallel columns, first column premises and the second one is responses and the directions for matching the two columns. Matching test items are also selection items specially used to measure a student's ability to identify the relationship between a set of similar items, each of which has two components. Such as words and their definitions, symbols and their meanings, dates and events, people and their accomplishments, etc. Matching exercise is economical methods when used with content which has the sufficient homogeneous factual information. In developing matching items, there are two columns of material. Matching exercise is used when measuring a student's ability to identify the relationship between a set of similar items, each of which has two components.

Q. Item Response theory (IRT)

—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 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.

Q. Define term NEAS?

The National Education Assessment System has been institutionalized in Pakistan at national level with the cooperation of provincial and area Assessment Centers. NEAS was established as five years development project with the financial assistance of the World Bank and Development for International Development (DfID) in year 2003. NEAS is subordinate office under the ministry of Federal Education & Professional Training.

Q. Objectives of NEAS

- 1) Informing policy: the extent to which geography and gender are linked to inequality in student performance.
- 2) Monitoring standards: How well the curricula are translated into knowledge and skills
- 3) Identifying correlation of achievement: the principal determinants of student performance
- 4) Directing teachers' efforts and raising students' achievement: assisting teachers to use data to improve student performance.

Q. Define short questions

There are two types of item formats which are included in the category of supply type: short answers and completion and fill in the blanks, under the range of objective test items. Both the short answers and completion items are forms of supply item that can be answered by a word, phrase, number or symbol. They only differ in form of presentation. Short answers are

presented in the form of direct questions while completion items are incomplete statements. Both of them can be answered with a word, phrase, number, or symbol. Such items are frequently used for measuring knowledge of terminology, specific facts, principles and procedures etc.

Q. Advantages of matching exercises

- The major objective of matching exercise is its compact form, which makes it possible to measure a large amount of related factual material in a relatively short time.
- Another advantage is ease of construction. Poor matching items can be rapidly constructed, but good matching item requires a high degree of skills.

Q. Names of test theories

1. classical test theory (CTT) and
2. item response theory (IRT)

Q. Define Item difficulty

CTT does not raise a complex theoretical model to relate an examinee's ability to the probability of success on a particular item. CTT collectively considers a pool of examinees and empirically examines their success rate on an item. This success rate of a particular pool of examinees on an item, well known as the p value of the item, is used as the index for the item difficulty. In CTT, the item difficulty index p (p value), is the proportion of examinees correct on an item, expresses item difficulty on an Item. Item difficulty in CTT is simply calculated by the percentage of students that correctly answered the item as refers to the p value which range from .00 to 1.00. The values closer to 1 more easy will be the item and conversely the values near to .00 the more difficult will be the item. The values lie somewhere in the middle i.e. 0.4 to 0.6 will refer to moderate item difficulty index.

Q. ICC slope 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

Q. Recording test items

The card contain information concerning the instructional objectives, specific objective, difficulty index, discrimination index and the content measured by the item should be prepared for each item to maintain its record.

Q. Probability of Guessing in Item Response Theory (IRT)

Guessing means giving an answer or making a judgment about something without being sure of all the facts. Guessing is a standard test-taking strategy presented to examinees taking a multiple choice assessment. If test scores are based simply on the number of questions answered correctly, then a random guess increases the chance of a higher score. In IRT this parameter of an item is also known as G (guessing) parameter which allow to detect the potential possibility of guessing in an item

Q. Item difficulty with reference to IRT

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. 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).

Q. difference b/w scoring criteria and scoring rubric

Scoring Criteria

Planning how responses will be scored leads to rethinking and clarifying the questions so that students have a clearer idea of what is expected. Clear specification of scoring criteria in advance of administering essay questions can contribute to improved reliability and validity of the assessment.

Scoring Rubric

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. Scores levels identified in a scoring rubric must be descriptive, not merely judgmental in nature. Example: Define the level of rubric as —Writing is clear and thoughts are completell as compared to —excellenll.

Q. What are the things that we measure in reading test

1. Vocabulary
2. Reading comprehension
3. Rate of reading

Q. 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.

Q. 5 limitations of matching exercises

- It is restricted to the measurement of factual information, based on rote learning.
- It is highly susceptible to the presence of irrelevant clues.
- Difficulty of findings homogenous material that is significant from the viewpoint of our objectives and learning outcomes.

Q. Learning outcomes in essay type Question

1. Analysis of relationship.
2. Compare and contrast positions.
3. Explain cause-effect relationship.
4. Organize data and support a viewpoint.
5. Formulate hypotheses.

Q. Advantages of Alternative Form (True-False)

True-false questions are well suited for testing lower level learning outcomes, like a student's ability to;

- i. Identify the correctness of factual statements, e.g. Earth is a planet.
- ii. Definition of terms e.g. Photosynthesis is the process by which leaves make food for plants.
- iii. Statement of principles, e.g. Earth is revolving around the sun.
- iv. Distinguish facts from opinion, e.g. Islam is the official religion of Pakistan.
- v. Recognize cause-and-effect relationship

Q. 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
3. Item banks are developed
4. Psychometric properties of item are tested
5. The process goes on round the year

Q. A typical item characteristic curve (ICC)

Three item characteristic curves are presented on the same graph. All have the same level of discrimination but differ with respect to difficulty.

The left-hand curve represents an easy item because the probability of correct response is high for low-ability examinees and approaches 1 for high-ability examinees.. The center curve represents an item of medium difficulty because the probability of correct response is

low at the lowest ability levels, around 0.5 in the middle of the ability scale and near 1 at the highest ability levels.

The right-hand curve represents a hard item. The probability of correct response is low for most of the ability scale and increases only when the higher ability levels are reached. Even at the highest ability level shown (+3), the probability of correct response is only 0.8 for the most difficult item

Q. Suggest how to write good short Questions ?

1. Word the item so that the required answer is both brief and specific.
2. Do not take statements directly from textbooks to use as a basis for short answer items.
3. A direct question is generally more desirable than an incomplete statement.
4. If the answer is to be expressed in numerical units, indicate the type of answer.
5. When completion items are used, do not include too many blanks or start with a blank space. Also avoid asking trivial information in blank space.

Q. Characteristics of Standardized Achievement Test

1. The test items are of highly technical quality. They have been developed by educational and test specialists, pretested and selected on the basis of difficulty discriminating power, and relationship to a clearly defined and rigid set of specifications.
2. Directions for administering and scoring are so precisely stated that the procedures are standard for different users of test.
3. Norms based on national samples of students in the grade where the test is intended for use are provided as aids in interpreting the test scores.
4. Equivalent and comparable forms of the test are usually provided, information concerning the degree to which the forms are comparable.
5. A test manual is used as guides for administering the test, evaluating its technical qualities, and interpreting and using the results.

Q. Recommendation for effective HST

- 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

Q. Approaches to holistic scoring

One way to implement holistic rubric is to decide beforehand on the number of quality categories into which you will sort the student's answers. A second better way of using the holistic method is to craft a holistic rubric, which defines the qualities of paper that belong in each category. For example, defining what an —A|| paper is a —B|| paper is and so on. A third refinement is to select specimen papers, which are good examples of each scoring category. Then you can compare the student's paper with the pre-specified specimens that define each category level.

A fourth way of implementing holistic rubric is to read the answer completely and one with another to decide which are the best. The next best and so on. This will result in the rough ranking of all the papers this approach of holistic rubric cannot be applicable to a large number of papers.

Among these four approaches first three are consistent with a grading philosophy of criterion referenced or absolute quality standards

While the fourth one is consistent with the norm reference or relative standard grading philosophy.

Q. Criterion referenced and norm referenced interpretation

Criterion-referenced Interpretation

In case of standardized test, interpretation can only be made with reference to the constructions on which test was based by the developer. This is primarily useful in mastery testing where a clearly defined and delimited domain of learning tasks can be most readily obtained. Such interpretations must be made with caution because these tests were typically designed to discriminate among individuals rather than describe the specific tasks they can perform. Criterion-referenced interpretations of test results are most meaningful when the test has been specifically designed for this purpose e.g. designing a test that measures a set of clearly stated learning tasks

Norm-referenced Interpretation

This interpretation tells us how an individual compares with other persons who have taken the same test e.g. ranking of scores from highest to lowest and to note where an individual's score falls. Standardized tests typically have been designed for norm-referenced interpretations which involves converting the raw scores to derived scores by means of table of norms.

Q. Note on matching exercise or Q. Limitation of matching exercise

Matching exercise is economical methods when used with content which has the sufficient homogeneous factual information. In developing matching items, there are two columns of

material. Matching exercise is used when measuring a student's ability to identify the relationship between a set of similar items, each of which has two components.

Advantages of Matching Exercises

- ✚ The major objective of matching exercise is its compact form, which makes it possible to measure a large amount of related factual material in a relatively short time.
- ✚ Another advantage is ease of construction. Poor matching items can be rapidly constructed, but good matching item requires a high degree of skills.

Limitations of Matching Exercises

- ✚ It is restricted to the measurement of factual information, based on rote learning.
- ✚ It is highly susceptible to the presence of irrelevant clues.
- ✚ Difficulty of findings homogenous material that is significant from the viewpoint of our objectives and learning outcomes.

Q. Advantages of essay type -5

Extended response questions have following advantages over other types of question formats: Effective for assessing higher order abilities: analyze, synthesize and evaluate. It is comparatively less time consuming to develop such items.

Emphasizes essential communications skills

Guessing is eliminated.

Q. Guessing in IRT-5

Guessing means giving an answer or making a judgment about something without being sure of all the facts. Guessing is a standard test-taking strategy presented to examinees taking a multiple choice assessment. If test scores are based simply on the number of questions answered correctly, then a random guess increases the chance of a higher score.

In IRT this parameter of an item is also known as G (guessing) parameter which allow to detect the potential possibility of guessing in an item.

There are two types of guessing

- ✚ Blind guessing
- ✚ Informed guessing

Blind guessing: Where an examinee chooses an answer at random from among the alternatives offered.

Informed guessing: Where the examinee draws upon all his knowledge and abilities to choose the answer most likely to be correct.

Item writers should be conscious of guessing and not write item that could be prone to guessing. IRT method of item analysis should be employed to eliminate those items prone to guessing

Q. Methodologies properties of ICC

1. The difficulty which under item response theory describes the item functions along the ability scale. For example an easy item functions among the low-ability examinees and a hard item functions among the high-ability examinees
2. The second property is discrimination, which describes how well an item can differentiate between examinees having abilities below the item location and those having abilities above the item location
3. An item characteristic curve is the graphical representation of the probability of answering an item correctly with the level of ability on the construct being measured.

Q. How can a teacher maximize achievement motivation

1. Encourage to do best, mitigate the fear
2. Highlight value of giving best
3. Reduce panic
4. Encourage serious thinking

Q. Difference between standardized tests and class test

- Quality of test items
- Reliability of the tests
- Procedure for administering and scoring
- Interpretation of scores
- The nature of the learning outcomes and the content measured