

INDEX

1:-<u>EVALUATION TECHNIQUES</u>.....	3
a) Observations	
b) Interviews	
c) Questionnaires	
d) Formative	
2:-What is Reliability? Describe its types in detail.....	4-5
3:-What is Statistics? Discuss the role of Statistics in Education and Research?	6
4:-<u>Characteristics of Good Measurement Instruments</u>.....	7
a) <u>Consistency (Reliability):</u>	7
b) <u>Meaningfulness (Validity):</u>	7-8
c) <u>Objectivity</u>	8
5:-<u>Principles of Evaluation</u>	9
a) Diagnostic Evaluation	
b) Formative Evaluation	
c) Placement Evaluation	
d) Summative Evaluation	
6:-<u>Measures of central tendency:</u>	10
<u>Advantage and Disadvantage of Mean, Median and Mode</u>	
7:-<u>Test Preparation Tips</u>	11
8:- <u>Administering the Test/ Trying Out</u>	12
9:- <u>Preparation for the Test</u>	12-13
10:-<u>Anecdotal Records</u>	14
11:-<u>Data collection</u>	15
12:-<u>Frequency distribution</u>.....	16
13:-<u>Graphical Representation of Data</u>	17
14:-<u>What are the classifications of data in statistics?</u>	18
15:- <u>WHAT IS A TEST</u>	19
TYPES OF TEST	
a) ESASY TYPE	
b) OBJECTIVE TYPE	

Assignment1

Q1:- Describe the **role of Measurement and Evaluation** in Education in detail. State the **principle of Evaluation**.....1-2

Q2:-Differentiate between the:

Formative and **Summative** Evaluation.....3

Placement and **Diagnostic** Evaluation.....4

Q3:- What are the different **Techniques of Evaluation**? Mention their merits and demerits.....5-6-7-8

Q4:-Differentiate between the:

a) **Anecdotal** and **Commulative** record.....9

b) **Interview** and **Questionnaire**.....9

c) **Rating Scales** and **Ranking Scale**.....10

Q5:-What are **Objective Type Tests**? Discuss their deferent **types** with examples.11

Assignment2

Q1:- What is **Reliability**? Describe its types in detail.....12-13-14-15

Q3:- What is **statistics**? Discuss the **role of Statistics** in Education and Research.16-17

Q4:- What do you know about **collection and classification of data**? Explain the "Editing and "Tabulation" process.....18-19-20

Q5:- Write notes on the following:

a) Validity.....21-22

b)Essay Type Test.....22

EVALUATION TECHNIQUES

There are several evaluation techniques or ways of collecting data. It is important for the evaluator to select the best technique to suit the purpose, and choose which programme leaders can be used to check the value of their programmes. These include tests, interviews, observations, questionnaires, formative, inspection of records, discussions and open box suggestions.

1. Observations

This is a visual technique where the evaluator observes, and records, any findings. Here, the evaluator decides whether he/she will observe as an outsider, or as a participant. This requires full concentration by the observer, who should study an aspect in detail, after consultation with other programme implementers. The observer actually serves as a second person for what other implementers are looking for. For this technique to be worthwhile, the observer should report and discuss any findings immediately, so that possible distortions, or failure to remember, do not affect the results. When using this technique, observers need to have sharp eyes and ears to be able to collect adequate data. The observer should carefully guard against bias.

2. Interviews

Interviews are a basic evaluation technique. They can be structured or unstructured. This requires a good ability to listen. A more structured interview would require the preparation of interview questions prior to the interview itself. It is important that data be recorded. The evaluator may use a tape recorder, in order to transcribe the relevant parts later, or take notes throughout the interview. Some evaluators depend on their memory, but this leads to a great risk of missing out important information. It is important that notes taken are discussed with the person interviewed, to check whether the information was correctly recorded.

3. Questionnaires

This involves a structured series of questions and statements that enable the evaluators to gather information about a particular programme. The advantage is that it can be sent to clients, to be completed in the absence of the evaluator. Unlike an interview, it also gives the evaluator an opportunity to cover a larger number of people at one time. All evaluation techniques require planning on the part of the evaluator. It is important for the evaluator to consult with relevant beneficiaries on the evaluation techniques they would like to use, so that conflicts of interest are avoided.

4. Formative

Formative evaluation, including pre-testing, is designed to assess the strengths and weaknesses or materials or campaign strategies before implementation. Formative research tailors the program to the target audience. Messages or products are tested by a small group before they are implemented on a large scale. This type of evaluation permits necessary revisions before the full effort goes forward. Its basic purpose is to maximize the change for program success before the activity starts.

Q: What is **Reliability**? Describe its types in detail.

'Reliability' of any research is the degree to which it gives an accurate score across a range of measurement. It can thus be viewed as being 'repeatability' or 'consistency'.

In summary:

- **Inter-rater**: Different people, same test.
- **Test-retest**: Same people, different times.
- **Parallel-forms**: Different people, same time, different test.
- **Internal consistency**: Different questions, same construct.

Inter-Rater Reliability

When multiple people are giving assessments of some kind or are the subjects of some test, then similar people should lead to the same resulting scores. It can be used to calibrate people, for example those being used as observers in an experiment.

Inter-rater reliability thus evaluates reliability across different people.

Two major ways in which inter-rater reliability is used are (a) testing how similarly people categorize items, and (b) how similarly people score items.

This is the best way of assessing reliability when you are using observation, as observer bias very easily creeps in. It does, however, assume you have multiple observers, which is not always the case.

Inter-rater reliability is also known as inter-observer reliability or inter-coder reliability.

Examples

Two people may be asked to categorize pictures of animals as being dogs or cats. A perfectly reliable result would be that they both classify the same pictures in the same way.

Test-Retest Reliability

An assessment or test of a person should give the same results whenever you apply the test.

Test-retest reliability evaluates reliability across time.

Reliability can vary with the many factors that affect how a person responds to the test, including their mood, interruptions, time of day, etc. A good test will largely cope with such factors and give relatively little variation. An unreliable test is highly sensitive to such factors and will give widely varying results, even if the person re-takes the same test half an hour later.

Generally speaking, the longer the delay between tests, the greater the likely variation. Better tests will give less retest variation with longer delays.

Of course the problem with test-retest is that people may have learned and that the second test is likely to give different results.

This method is particularly used in experiments that use a no-treatment control group that is measure pre-test and post-test.

Examples

Various questions for a personality test are tried out with a class of students over several years. This helps the researcher determine those questions and combinations that have better reliability.

In the development of national school tests, a class of children are given several tests that are intended to assess the same abilities. A week and a month later, they are given the same tests. With allowances for learning, the variation in the test and retest results are used to assess which tests have better test-retest reliability.

Parallel-Forms Reliability

One problem with questions or assessments is knowing what questions are the best ones to ask. A way of discovering this is do two tests in parallel, using different questions.

Parallel-forms reliability evaluates different questions and question sets that seek to assess the same construct.

Parallel-Forms evaluation may be done in combination with other methods, such as Split-half, which divides items that measure the same construct into two tests and applies them to the same group of people.

Examples

An experimenter develops a large set of questions. They split these into two and administer them each to a randomly-selected half of a target sample.

In development of national tests, two different tests are simultaneously used in trials. The test that gives the most consistent results is used, whilst the other (provided it is sufficiently consistent) is used as a backup.

Internal Consistency Reliability

When asking questions in research, the purpose is to assess the response against a given construct or idea. Different questions that test the same construct should give consistent results.

Internal consistency reliability evaluates individual questions in comparison with one another for their ability to give consistently appropriate results.

Average inter-item correlation compares correlations between all pairs of questions that test the same construct by calculating the mean of all paired correlations.

Average item total correlation takes the average inter-item correlations and calculates a total score for each item, then averages these.

Split-half correlation divides items that measure the same construct into two tests, which are applied to the same group of people, and then calculates the correlation between the two total scores.

Q: - What is **Statistics**? Discuss the role of Statistics in Education and Research?

Statistics is the science of collecting, analyzing and making inference from data. Statistics is a particularly useful branch of mathematics that is not only studied theoretically by advanced mathematicians but one that is used by researchers in many fields to organize, analyze, and summarize data. Statistical methods and analyses are often used to communicate research findings and to support hypotheses and give credibility to research methodology and conclusions. It is important for researchers and also consumers of research to understand statistics so that they can be informed, evaluate the credibility and usefulness of information, and make appropriate decisions

The role of statistics in the development of quality in education

Statistics are a potent and effective tool for the development of quality in education. As education systems become increasingly complex, so also is the availability of accurate, up-to-date statistics a clear imperative. This by no way suggests that statistics on their own ensure or guarantee quality, but rather that their availability are a condition for achieving it. The role of statistics in the promotion of quality can be examined by linking it to each of the stages of the model presented in Figure 1. Alternatively, statistics can be viewed as part of the broader process of quality assurance that includes educational planning, policy formulation, management, monitoring and evaluation of the education systems, and highlighting the role played by statistics in strengthening the process itself.

The current situation

Slightly over a decade ago many countries in Sub Saharan Africa (SSA) had serious capacity gaps in the generation and development of statistical information. This was a time when there was great need to make informed decisions on the judicious allocation of limited resources. Investment in building information capacity did not increase in accordance with the demand for information. Instead the capacity for providing the requisite information deteriorated further, rendering decision-makers less capable of making informed decisions. The majority of African countries in SSA had very little or no up-to-date statistical information on education. The pages on SSA in the UNESCO Year-book were largely blank and the education policies of Africa lacked the quantitative dimension, although their goals were expressed in quantitative terms. So the goals and targets of the countries were not anchored to the reality of the situation. Statistical capacity building process

- 1 Lacking statistical infrastructure; little government commitment and use of data; less need for international comparable data
- 2 Basic data channels in place; some commitment to data use; Data fragments across ministries; coverage and relevance; regional comparisons
- 3 Stable information Systems; good links between users and producers of data; responsive to relevant policy issues but the demands are more complex. International comparisons used widely.

Characteristics of Good Measurement Instruments

Measurement tools can be judged on a variety of merits. These include practical issues as well as technical ones. All instruments have strengths and weaknesses—no instrument is perfect for every task. Some of the practical issues that need to be considered include:

Cost

Availability

Training required

Ease of administration, scoring, analysis

Time and effort required for respondent to complete measure

Along with the practical issues, measurement tools (especially surveys, tests and scales) may be judged on the following technical characteristics:

1) Consistency (Reliability): A good instrument will produce consistent scores. An instrument's reliability is estimated using a correlation coefficient of one type or another. For purposes of service learning research, the major characteristics of good scales include:

1 Temporal consistency—the ability of an instrument to give accurate scores from one time to another. Also known as test-retest reliability.

1 Coherence—the consistency of items within a test. There are two types of item coherence: split-half reliability assesses the consistency of items in one-half of a scale to the other half. Internal consistency reliability estimates the consistency among all items in the instrument (typically measured using Chronbach's alpha coefficient.)

1 Scoring agreement—the degree to which different observers or raters give consistent scores using the same instrument, rating scale, or rubric. Also called inter-rater reliability.

2.) Meaningfulness (Validity): A valid measurement tool does a good job of measuring the concept that it purports to measure. It is important to remember that the validity of an instrument only applies to a specific purpose with a specific group of people. For example, a scale is not considered simply "valid" or "invalid"—but it might be considered valid for measuring social responsibility outcomes with college freshmen. There are three main classes of validity

Construct validity— The concept (e.g., intelligence, moral development, content knowledge) that is being measured is sometimes called the "construct." Construct validity establishes that the instrument is truly measuring the desired construct. This is the most important form of validity, because it really subsumes all .of the other forms of validity

1. Convergent validity
2. Discriminant validity
3. Factor structure

Content validity—Establishes that the instrument includes items that comprise the relevant content domain. (For example, a test of English grammar might include questions on subject-verb agreement, but should not include items that test algebra skills) 1

1. Face validity

Criterion-related validity— The instrument “behaves” the way it should given your theory about the construct

1. Concurrent validity

2. Predictive validity

Objectivity

An objective data collection tool provides measurements that are uninfluenced or undistorted by the beliefs or biases of the researcher who applies it. Most instrumentation provides objective data (e.g., a patient's measured height is not likely to be distorted by the researcher). However, data collected by observation, surveys or tests can lack objectivity (e.g., differences in interpretation of human behaviors, essay exam grading, etc).

Principles of Evaluation

Evaluation should be

1. Based on clearly stated objectives
2. Comprehensive
3. Cooperative
4. Used Judiciously
5. Continuous and integral part of the teaching – learning process

Types of Evaluation used in classroom instruction

1. **Diagnostic Evaluation** – detects pupil's learning difficulties which somehow are not revealed by formative tests. It is more comprehensive and specific.
2. **Formative Evaluation** – It provides feedback regarding the student's performance in attaining instructional objectives. It identifies learning errors that needed to be corrected and it provides information to make instruction more effective.
3. **Placement Evaluation** – It defines student's entry behaviors. It determines knowledge and skills he possesses which are necessary at the beginning of instruction.
4. **Summative Evaluation** – It determines the extent to which objectives of instruction have been attained and is used for assigning grades/marks and to provide feedback to students.

Measures of central tendency

Advantage and Disadvantage of Mean, Median and Mode

MEAN – The most widely used and familiar average. The most reliable and the most stable of all measures of central tendency.

Advantage: It is the best measure for regular distribution.

Disadvantage: It is affected by extreme values

MEDIAN – The scores that divides the distribution into halves. It is sometimes called the counting average.

Advantage: It is the best measure when the distribution is irregular or skewed. It can be located in an open-ended distribution or when the data is incomplete (ex. 80% of the cases is reported)

Disadvantage: It necessitates arranging of items according to size before it can be computed

MODE – the crude or inspectional average measure. It is most frequently occurring score. It is the poorest measure of central tendency.

Advantage: Mode is always a real value since it does not fall on zero. It is simple to approximate by observation for small cases. It does not necessitate arrangement of values.

Disadvantage: It is not rigidly defined and is inapplicable to irregular distribution

Test Preparation Tips

- ✓ Preparation for your first test should begin on the first day of class; this includes paying attention during class, taking good notes, studying, completing homework assignments and reviewing study materials on a regular basis.

- ✓ Budget your time, make sure you have sufficient time to study so that you are well prepared for the test.

- ✓ Go to review sessions, pay attention to hints that the instructor may give about the test. Take notes and ask questions about items you may be confused about.

- ✓ Ask the instructor to specify the areas that will be emphasized on the test.

- ✓ Make sure you go to the class right before the test; it's another prime time for the instructor to give out more hints or the format of the test.

- ✓ Go over any material from practice tests, HW's, sample problems, review material, the textbook, class notes...

- ✓ Eat before a test. Having food in your stomach will give you energy and help you focus but avoid heavy foods which can make you groggy.

- ✓ Don't try to pull an all nighter. Get at least 3 hours of sleep before the test (normally 8 hours of sleep a night is recommended but if you are short on time, get at least 3 hours so that you'll be well rested enough to focus during the test).

- ✓ Put the main ideas/information/formulas onto a sheet that can be quickly reviewed many times, this makes it easier to retain the key concepts that will be on the test.

- ✓ Try to show up at least 5 minutes before the test will start.

- ✓ Set your alarm and have a backup alarm set as well.

- ✓ Go to the bathroom before walking into the exam room. You don't want to waste anytime worrying about your bodily needs during the test.

Administering the Test/ Trying Out

Introduction

Administering the written test is perhaps the most important aspect of the examining process. The atmosphere the test administrator creates in the test room and the attitude the test administrator displays in performing his/her duties is extremely important. The test administrator's manner, bearing, and attitude may well inspire confidence in competitors and put them at ease while participating in the testing process.

Test administrators should make certain that they and their monitors are well prepared to administer the test. There is an understandable tendency on the part of some test administrators to develop their own style of conducting tests because they have conducted the same test many times and are certain that they can do it correctly. While we encourage the creation of a personal style in establishing a good relationship with competitors, we consider it equally important that test administrators not deviate from the directions for conducting the test. The following material should serve as a reminder of decisions to be made before, during, and after each test. In addition, it should eliminate confusion and uncertainty that sometimes arises when administering tests.

Preparation for the test

It is important for test administrators to understand the nature and purpose of the particular test. What type of test is being given? What job(s) is it used to fill? What is the reason for giving this particular test? Is it a regularly scheduled or a special test situation?

To the extent possible, test administrators should:

Prepare...	Why Important...
Room Arrangement	<ul style="list-style-type: none"> To ensure that the room is large enough to accommodate all competitors comfortably with adequate light and ventilation and that the noise level will be such that competitors will not be distracted. Seating arrangements are such that all competitors will be able to see any instructions
Scheduling of test	<ul style="list-style-type: none"> To ensure that there is no conflict with other scheduled tests or with other functions- the test must not be interrupted. Monitors know they are to be in the examination room in advance of the scheduled time. A minimum of 30 minutes is recommended; for some situations where more specific advance preparations are required, more time will be needed
Test monitors	<ul style="list-style-type: none"> Arrangements should be made for the proper number of test monitors. There should be one test monitor for every 30 competitors. The ratio of test administrators/monitors should be as follows: <p>Competitors, Test Administrators, Monitors 1 - 30, 1, 0 31 - 60, 1, 1 61 - 90, 1, 2 91 - 120, 1, 3</p> <p>Since only OPM trained test administrators may conduct tests, it is advisable to have an alternate, OPM trained test administrator, available in the event the scheduled test administrator is unable to conduct the test as planned.</p>
Order test materials	<ul style="list-style-type: none"> The test administrator should be certain that they have the correct tests and related materials, including the Directions for Conducting

	<p>the Test (DFC), # 2 pencils, a timer (or stopwatch), a "Testing--Do Not Disturb" sign, and other supplies (scratch paper, etc.) as required by the DFC. Test administrators should count the number of test booklets at least twice to verify that they have received the correct number.</p>
<p>Study directions for conducting test</p>	<p>The test administrator should study the Directions for Conducting the Test (DFC) carefully before the examination. He/she should be familiar with:</p> <ul style="list-style-type: none"> • The general make-up of the test; • Time limits involved; • Special directions; • Method of indicating answers; • Any sample questions (if they are to be answered in the test room); and <p>Any changes to the instructions in the DFC. These instructions will be provided by the appropriate OPM Services Branch.</p>
<p>Review with monitors how materials are to be handled</p>	<p>The test administrator should review with the monitors the order in which materials are to be handed out and collected. He/she should also discuss how doors of the room are to be covered during entrance by competitors before the test and exit following the test.</p> <p>(Note: Advance preparation can help preclude potential for compromise of test material by competitors leaving through unattended doors, taking test material with them - one of the more common reasons for reported test material losses).</p>

Anecdotal Records

An **anecdote** is an account of an event in a child's day. The record of this event can be detailed or brief. These short reports describe, in a factual way, the incident, its context, and what was said or done by the participant(s). In most cases, anecdotes focus on very simple, everyday interactions among children, children and adults as well as children and materials in the environment.

Ideally, the anecdotal record should be recorded as it unfolds or immediately after. However, anecdotal records usually have to be written later at the end of the day. Keeping brief notes on index cards or sticky notes carried in your pockets can be helpful. Jotting one-word reminders or short phrases on the cards about the event can provide a set of reminders when the anecdote is written.

Characteristics of Anecdotal Records:

- Simple reports of behavior'
- Result of direct observation.
- Accurate and specific
- Gives context of child's behavior
- Records typical or unusual behaviors

Purpose:

Anecdotes capture the richness and complexity of the moment as children interact with one another and with materials. These records of child behavior and learning accumulated over time enhance the teacher's understanding of the individual child as patterns or profiles begin to emerge. Behavior change can be tracked and documented, and placed in the child's portfolio resulting in suggestions for future observations, curriculum planning and student or parent conferences.

Examples of Anecdotal Records:

Child's Name: Melissa C.

Date & Time: 4/23/01 8:45 am

Place or Learning Center: Preschool classroom-free play

Data collection

Data collection is a term used to describe a process of preparing and collecting [data](#), for example, as part of a [process improvement](#) or similar project. The purpose of data collection is to obtain information to keep on record, to make decisions about important issues, or to pass information on to others. Data are primarily collected to provide information regarding a specific topic.^[1]

Data collection usually takes place early on in an improvement project, and is often formalised through a [data collection plan](#)^[2] which often contains the following activity.

1. Pre collection activity — agree on goals, target data, definitions, methods
2. Collection — data collection
3. Present Findings — usually involves some form of sorting^[3] analysis and/or presentation.

Prior to any data collection, pre-collection activity is one of the most crucial steps in the process. It is often discovered too late that the value of their interview information is discounted as a consequence of poor sampling of both questions and informants and poor elicitation techniques.^[4] After pre-collection activity is fully completed, data collection in the field, whether by interviewing or other methods, can be carried out in a structured, systematic and scientific way.

A formal data collection process is necessary as it ensures that data gathered are both defined and accurate and that subsequent decisions based on arguments embodied in the findings are valid.^[5] The process provides both a baseline from which to measure from and in certain cases a target on what to improve.

Other main types of collection include census, sample survey, and administrative by-product and each with their respective advantages and disadvantages. A census refers to data collection about everyone or everything in a group or population and has advantages, such as accuracy and detail and disadvantages, such as cost and time. A sample survey is a data collection method that includes only part of the total population and has advantages, such as cost and time and disadvantages, such as accuracy and detail. Administrative by-product data are collected as a byproduct of an organization's day-to-day operations and has advantages, such as accuracy, time simplicity and disadvantages, such as no flexibility and lack of control.^[6]

Frequency distribution

In [statistics](#), a **frequency distribution** is an arrangement of the values that one or more variables take in a [sample](#). Each entry in the table contains the [frequency](#) or count of the occurrences of values within a particular group or interval, and in this way, the table summarizes the [distribution](#) of values in the sample.

Example: Goals

Sam's team has scored the following numbers of goals in recent games: 2, 3, 1, 2, 1, 3, 2, 3, 4, 5, 4, 2, 2, 3

Sam put the numbers in order,
then added up:

- how often 1 occurs (2 times),
- how often 2 occurs (5 times), [etc](#), and wrote them down as a Frequency Distribution table From the table we can see interesting things such as

- getting 2 goals happens most frequently
- only once did they get 5 goals

Joint frequency distributions

Bivariate joint frequency distributions are often presented as (two-way) [contingency tables](#):

<i>Two-way contingency table with marginal frequencies</i>				
	Dance	Sports	TV	Total
Men	2	10	8	20
Women	16	6	8	30
Total	18	16	16	50

The *total* row and *total* column report the **marginal frequencies** or [marginal distribution](#), while the body of the table reports the **joint frequencies**.^[1]

Graphical Representation of Data

Whenever verbal problems involving a certain situation is presented visually before the learners, it makes easier for the learner to understand the problem and attempt its solution.

Similarly, when the data are presented pictorially (or graphically) before the learners, it makes the presentation eye-catching and more intelligible. The learners can easily see the salient features of the data and interpret them.

There are many forms of representing data graphically. They are

- (i) Bar graphs
- (ii) Histograms
- (iii) Frequency polygons
- (iv) Ogive
- (v) Pictographs
- (vi) Pie charts

In this lesson, we shall learn to read and draw Bar graphs, Histograms and Frequency polygons, and graphs related to day-to-day use, like temperature-time graph, velocity-time graph, pressure-volume graph, etc. Other forms of graphs are beyond the scope of the present lesson.

OBJECTIVES

After studying this lesson, the learner will be able to

- draw bar charts for given data
- draw a histogram and frequency polygon for given data
- read and interpret given bar charts and histograms
- read the relevant information from graphs relating to day-to-day activities, like
 - (i) temperature-time graph
 - (ii) velocity-time graph
 - (iii) Pressure volume graph, etc.

Draw graph relating to day-to-day activities, like the ones above.

28.3 EXPECTED BACKGROUND KNOWLEDGE

- Knowledge of drawing and marking axes
- Knowledge of drawing rectangles and plotting points
 - (ii) Practice of reading graphs.

What are the classifications of data in statistics?

Classifications of data

A. According to Nature

1. Quantitative data- information obtained from numeral variables(e.g. age, bills, etc)
2. Qualitative Data- information obtained from variables in the form of categories, characteristics names or labels or alphanumeric variables (e.g. birthdays, gender etc.)

B. According to Source

1. Primary data- first- hand information (e.g. autobiography, financial statement)
2. Secondary data- second-hand information (e.g. . biography , weather forecast from news papers)

C. According to Measurement

1. Discrete data- countable numerical observation.
 - Whole numbers only
 - has an equal whole number interval
 - obtained through counting(e.g. corporate stocks, etc.)
2. Continuous data-measurable observations.
 - decimals or fractions
 - obtained through measuring(e.g. bank deposits, volume of liquid etc.)

D. According to Arrangement

1. Ungrouped data- raw data
 - no specific arrangement
2. Grouped Data - organized set of data
 - at least 2 groups involved
 - arranged

Easy Type Test

WHAT IS A TEST

- TEST IS A SYSTEMATIC PROCEDURE FOR OBSERVING PERSONS AND DESCRIBING THEM WITH EITHER A NUMERICAL SCALE OR A CATEGORY SYSTEM. THUS TEST MAY GIVE EITHER A QUALITATIVE OR QUANTITATIVE INFORMATION

CONTENT

- TEST COMMONLY REFERS TO A SET OF ITEMS OR QUESTIONS UNDER SPECIFIC CONDITIONS.

TYPES OF TEST

- ESASY TYPE
- OBJECTIVE TYPE

ESSAY TYPE TEST

- IT IS AN ITEM FORMAT THAT REQUIRES THE STUDENT TO STRUCTURE A RATHER LONG WRITTEN RESPONSE UP TO SEVERAL PARAGRAPHS

CHARACTERISTICS OF ESSAY TEST

- GENERALLY ESSAY TESTS CONTAIN MORE THAN ONE QUESTION IN THE TEST
- ESSAY TESTS ARE TO BE ANSWERED IN WRITING ONLY
- ESSAY TEST TESTS REQUIRE COMPLETELY LONG ANSWERS
- ESSAY TESTS ARE ATTEMPTED ON THE BASIS OF RECALLING THE MEMORY

TYPES OF ESSAY TEST

SELECTIVE RECALL (BASIS GIVEN)

- EVALUATION RECALL (BASIS GIVEN)
- COMPARISON OF TWO THINGS ON A SINGLE DESIGNATED BASIS
- COMPARISON OF TWO THINGS IN GENERAL
- DECISIONS (FOR AND AGAINST)
- EXPLANATION OF THE USE EXACT MEANING OF SOME WORD, PHRASE OR STATEMENT

CONTENT

- SUMMARY OF SOME UNIT OF THE TEXT OR OF SOME ARTICLE
- ANALYSIS
- ILLUSTRATIONS OR EXAMPLES
- APPLICATION OF RULES, LAWS, OR PRINCIPLES TO NEW SITUATIONS
- DISCUSSIONS
- CRITICISM
- INFRENTIAL THINKING

ADVANTAGES

- CAN MEASURE COMPLEX LEARNING OUTCOMES
- EMPAHASIZE INTEGRATION AND APPLICATION OF THINKING AND PROBLEM SOLVING
- CAN BE EASILY CONSRUCTED
- EXAMINEE FREE TO RESPOND
- NO GUESSING AS IN OBJECTIVE ITEM

CONTENT

- REQUIRE LESS TIME FOR TYPING, DUPLICATING OR PRINTING, CAN BE WRITTEN ON BOARD
- CAN BE USED AS DEVICE FOR MEASURING AND IMPROVING LANGUAGE AND EXPRESSION SKILLS

LIMITATIONS

LACK OF CONSISTENCY IN JUDGEMENTS EVEN AMONG COMPETENT EXAMINERS

- THEY HAVE HOLO EFFECT
- QUESTION TO QUESTION CARRY EFFECT
- EXAMINEE TO EXAMINEE CARRY EFFECT
- LANGUAGE MECHANIC EFFECT
- LIMITED CONTENT VALIDITY

CONTENT

- SOME EXAMINERS ARE TOO STRICT AND SOME ARE TOO LINIENT
- DIFFICULT TO SCORE OBJECTIVELY
- TIME CONSUMING
- LANGTHY ENUMERATION OF MEMORIZED FACTS

SUGGESTIONS FOR CONSRUCTION OF EASSY TESTS

- ASK QUESTIONS THAT REQUIRE THE EXAMINEE TO SHOW COMMAND OF ESSENTIAL KNOWLEDGE
- MAKE QUESTIONS AS EXPLICIT AS POSSIBLE
- SHOULD BE NO CHOICE IN QUESTIONIN QUESTION PAPER
- TEST CONSTRUCTOR SHOULD PREPARE IDEAL ANSWERS TO ALL QUESTIONS
- INTIMATE THE EXAMINEE ABOUT DESIRED LANGTH OF THE ANSWERS

CONTENT

- MAKE EACH QUESTION RELATIVELY SHORT BUT INCREASE NUMBER OF QUESTIONS
- TEST CONSTRUCTOR SHOULD GET HIS TEST REVIEWD Y ONE AO MORE COLLEAGUES
- QUESTIONS SHOULD BE SO WORDED THAT ALL EXAMINEES INTERPRET THEM IN THE SAME WAY AS THE EXAMINER WANTS

SUGGESTIONS FOR SCORING ESSAY TESTS

- PREPARE SCORING GUIDE IN THE FORM OF OUTLINE
- PARTICULAR QUESTION SHOULD BE SCORED AT ONE TIME OF ALL THE EXAMINEES
- TO AVOID HOLO EFFECT, IDENTITY OF THE EXAMINEE SHOULD NOT BE COMMUNICATED TO THE EXAMINER
- IF POSSIBLE APPOINT MORE THAN ONE EXAMINER. THE EXAMINERS SHOULD NOT KNOW WHO IS THE OTHER EXAMINER

CONTENT

- THE CORRECTNESS OF THE SUBJECT MATTER SHOULD NOT BE MIXED WITH THE GOOD HANDWRITING, BETTER LANGUAGE, IF THEY ARE TO BE GIVEN ANY WEIGHT, IT SHOULD BE CLEARLY INDICATED

SHORT ANSWER TYPE TESTS

- SHORT ANSWER ITEMS REQUIRES THE EXAMINEE TO RESPOND TO THE ITEM WITH A WORD, SHORT PHRASE, NUMBER OR A SYMBOL.

CHARACTERISTICS

- THE TEST HAS SUPPLY RESPONSE RATHER THAN SELECT OR IDENTIFY
- IN THE FORM OF QUESTION OR INCOMPLETE STATEMENT
- THE TEST CAN BE ANSWERED BY A WORD, A PHRASE, A NUMBER OR SYMBOL

FORMS OF SHORT ANSWER ITEMS

- QUESTION FORM
- IDENTIFICATION OR ASSOCIATION FORM
- COMPLETION FORM

ADVANTAGES

- VERY EASY TO CONSTRUCT
- LOW PROBABILITY OF GUESSING THE ANSWER BECAUSE IT HAS TO BE SUPPLIED BY THE EXAMINEES RATHER THAN SELECT IDENTIFY FROM THE GIVEN ANSWERS
- THEY ARE GOOD TO TEST THE LOWEST LEVEL OF COGNITIVETAXONOMY (KNOWLEDGE, TERMINOLOGY, FACTS)

LIMITATIONS

- THEY ARE UNSUITABLE FOR MEASURING COMPLEX LEARNING OUTCOMES

SUGGESTIONS FOR CONSTRUCTION OF SHORT ANSWER TESTS

- AS FOR AS POSSIBLE QUESTION FORM SHOULD BE USED
- THE QUESTION SHOULD NOT BE PICKED UP EXACTLY FROM THE BOOK
- THE QUESTION SHOULD NOT PROVIDE ANY CLUE

CONTENT

- THE SCORING KEY SHOULD BE PREPARED
- THE BLANK SPACE IS TO BE COMPLETED BY AN IMPORTANT WORD RATHER THAN TREVIAL WORDS

OBJECTIVE TYPE TESTS

- ANY TEST HAVING CLEAR AND UNAMBIGIOUS SCORING CRITERIA
- TEST THAT CAN BE OBJECTIVELY SCORED

CHARACTERISTICS

- THEY CAN BE RELIABLY SCORED
- THEY ALLOW FOR ADEQUATE CONTENT SAMPLING

FORMS OF OBJECTIVE TYPE TESTS

A)TWO CHOICE ITEMS

1. TRUE/FALSE ITEMS
2. COMPLETION TYPE (IF TWO CHOICES ARE GIVEN AGAINST EACH BLANK)

B) MORE THAN TWO CHOICE ITEMS

1. MATCHING ITEMS
2. MCQs

TRUE/FALSE TESTS (SHOOTING QUESTIONS)

A TRUE FALSE ITEM CONSISTS OF A STATEMENT OR PROPOSITION WHICH THE EXAMINEE MUST JUDGE AND MARK AS EITHER TRUE OR FALSE

ADVANTAGES

- IT TAKES LESS TIME TO CONSTRUCT TRUE FALSE ITEMS
- HIGH DEGREE OF OBJECTIVITY
- TEACHER CAN EXAMINE STUDENTS ON MORE MATERIAL

LIMITATIONS

- HIGH DEGREE OF GUESING
- LARGELY LIMITED TO LEARNIGNOUTCOMES IN THE KNOWLEDGE AREA
- THEY EXPOSE STUDENTS TO ERROR WHICH IS PSYCHOLOGICALLY UNDESIREABLE
- THEY MAY ENCOURAGE STUDENTS TO STUDY AND ACCEPT ONLY OVERSIMPLIFIED STATEMENTS OF TRUTH ANS FACTUAL LEARNING

SUGGESTIONS

- BALANCE BETWEEN TRUE AND FALSE ITEMS
- EACH STATEMENT SHOULD BE UNEQUIVOCALLY TRUE OR FALSE. IT SHOULD NOT BE PARTLY TRUE OR PARTLY FALSE
- DOUBLE NEGATIVES SHOULD BE AVOIDED
- LONG AND COMPLEX STATEMENTS SHOULD NOT BE USED AS THEY MEASURE READING COMPREHENSION

CONTENT

- ONLY ONE IDEA SHOULD BE MEASURED IN ONE STATEMENT
- EXPLAIN WHICH JUDGEMENT IS TO BE USED TRUE/FALSE, YES/NO, CORRECT/INCORRECT
- CLUES SHOULD BE AVOIDED
- STATEMENTS SHOULD NOT BE TAKEN DIRECTLY FROM THE TEXTBOOK

MATCHING TYPE TESTS

A TEST CONSISTING OF A TWO COLUMN FORMAT, PREMISES AND RESPONSES THAT REQUIRES THE STUDENT TO TAKE A CORRESPONDENCE BETWEEN THE TWO

ADVANTAGES

- SIMLE TO CONSTRCUT AND SCORE
- WELL SUITED TO MEASURE ASSOCIATION
- REDUCE THE EFFECT OF GUESSING
- THEY CAN BE USED TO EVALUATE EXAMINEE'S UNDERSTANDING OF CONCEPTS, PRINCIPLE, SCHEMES FOR CLASSIFYNG OBJECTS, IDEAS OR EVENTS

LIMITATIONS

- THEY GENERALLY PROVIDE CLUES
- THEY ARE RESTRICTED TO FACTUAL INFORMATION WHICH ENCOURAGES MEMORIZATION
- IF THE SAME NUMBER OF ITEMS ARE WRITTEN IN BOTH THE COLUMNS, THE MATCHING TYPE IS CONVERTED TO MCQs AT LATE STAGE AND IN THE END IT IS CONVERTED TO TRUE AND FALSE CATEGORY

SUGGESTIONS

- HOMOGENEOUS ITEMS SHOLD BE SELECTED
- NO CLUE SHOULD BE PROVIDED IN BOTH THE COLUMNS
- CLEAR INSTRUCTION TO ATTEMPT
- ALL THE ITEMS SHOULD BE PRINTED ON THE SAME PAGE
- PREMISE SHOULD BE WRITTEN IN THE LEFT HAND COLUMNAMD BE NUMBERED, RESPONSES SHOULD BE WRITTEN IN THE RIGHT HAND COLUMN AND BE LETTERED

CONTENT

- RESPONSES SHOULD BE MORE THAN THE PREMISES TO ENSURE THAT EXAMINEE HAS TO THINK EVEN UPTO LAST PREMISE
- CLEAR DIRECTIONS
- INCOMPLETE SENTENCES SHOULD NOT BE USED FOR PREMISE

MULTIPLE CHOICE ITEMS

MULTIPLE CHOICE ITEMS CONSIST OF TWO PARTS: A STEM AND NUMBER OF OPTIONS OR ALTERNATIVES. THE STEM IS A QUESTION OR STATEMENT THAT IS ANSWERED OR COMPLETED BY ONE OF THE ALTERNATIVES. ALL INCORRECT OR LESS APPROPRIATE ALTERNATIVES ARE CALLED DISTRACTORS OR FOILS AND THE STUDENT'S TASK IS TO SELECT THE CORRECT OR BEST ALTERNATIVE FROM ALL THE OPTIONS.

FORMS OF MCQs

- 1.THE CORRECT ANSWER FORM
- 2.THE BEST ANSWER FORM
- 3.THE INCOMPLETE STATEMENT FORM
- 4.THE NEGATIVE FORM
- 5.THE COMBINED RESPONSE FORM
- 6.SUBSTITUTION FORM

ADVANTAGES

- THEY CAN MEASURE COMPLEX LEVEL KNOWLEDGE i.e. UNDERSTANDING, JUDGEMENT AND ABILITY TO SOLVE PROBLEMS
- A SUBSTANTIAL AMOUNT OF COURSE SONTENT CAN BE TESTED BECAUSE THE EXAMINEES DO NOT REQUIRE MUCH TIME FOR WRITING THE ANSWER
- OBJECTIVITY IN SCORING EVEN A LAYMAN CAN SCORE
- THEY CAN CHECK DISCRIMINATION ABILITY OF STUDENTS

CONTENT

- REDUCE THE EFFECT OF GUESSING
- CAN BE EASILY ADAPTED FOR MACHINE SCORING
- THIS FORMAT IS HELPFUL IN ITEM ANALYSIS

LIMITATIONS

- THEY REQUIRE EXAMINEE TO SELECT THE ANSWER FROM A FIXED LIST AND NOT PERMIT TO CREATE OR EXPRESS AND ORGANIZE THEIR OWN IDEAS
- EXAMINEES WHO DO NOT THE CORRECT ANSWER CAN SUCCEED IN GUESSING
- SOMETIMES DIFFICULT TO FIND FOUR OR FIVE CHOICES IN CONSTRUCTION
- TIME CONSUMING AT TEAHCER END

SUGGESTIONS

- STEM SHOULD INTRODUCE WHAT IS EXPECTED OF THE EXAMINEE
- SPECIFIC DETERMONERS SHOULD BE AVOIDED
- VOCABULARY ACCORDING TO THE LEVEL OF STUDENTS
- ALL THE CHOICES SHOULD BE PLAUSIBLE
- TEST ITEMS SHOULD HAVE DEFENSIBLE CORRECT OR BEST ANSWER
- THE CORRECT CHOICE SHOULD NOT BE AT THE SAME PLACE IN ALL OR MOST OF THE ITEMS

CONT...

- THE CHOICE LIKE "NONE OF THE ABOVE" "ALL THE ABOVE" SHOULD BE AVOIDED
- EACH ITEM SHOULD POSE ONLY ONE PROBLEM
- TEACHER SHOULD CONSTRUCT MCQs ON DAILY BASIS