

Educational measurement, assessment and evaluation

1. EDUCATIONAL MEASUREMENT, ASSESSMENT AND EVALUATION A Study Guide Approach
CREATED FOR BY: BOYET B. ALUAN
2. BASIC CONCEPTS AND PRINCIPLES A. TEST- B. MEASUREMENTS- C. ASSESSMENT D. EVALUATION-
3. TEST--- most commonly used method of making measurements in education. It is also an instrument or systematic procedures for measuring sample of behavior by posing a set of questioning a uniform manner. Designed to measure any quality, ability, skill or knowledge. There is right or wrong answer.
4. MEASUREMENT-- assignment of numbers (quantity), uses variety of instrument: test, rating scale. It is the process of obtaining numerical description of the degree of individual possesses. Quantifying of how much does learner learned.
5. ASSESSMENT--- process by which evidence of student achievement is obtained and evaluated. Information is to objective it include testing, interpreting and placing information in context. It is the process of gathering and organizing data—the basis for decision making (evaluation). Methods of measuring and evaluating the nature of the learner/(what he learned, how he learned).
6. EVALUATION--it is a process because it includes a series of steps (*establishing objective, *classifying objective, *defining objective, *selecting indicators, *and comparing data with objectives). It is concerned with making judgments on the worth or value of a performance, answers the question —how good, adequate, or desirable. It is also the process of obtaining, analyzing and interpreting information to determine the extent to which students achieve instructional objective.
7. HISTORY DEVELOPMENT OF INTELLIGENCE TEST
GERMAN
1. Ebbinghaus— quantitative memory
2. Kroeplin and Sommer— association test
3. William Stern— mental quotient $MQ = MA/CA$
4. Terman— Intelligent quotient $IQ = (MA/CA) * 100$
ENGLAND
1. Galton— questioning method and theory of eugenics statistical and experimental method
2. Pearson— method of correlation (Pearson- product- moment coefficient of correlation)
3. Spearman— rank correlation or Spearman Rho
4. Spearman-brown formula—reliability of full

test.5. Split-half method—half of the odd and even

8. FRANCE1. Esquirol—mental disability and insanity2. Sequin— board for mentally defectiveness using ten different sizes and shapes to be inserted in the hole.3. Binet— extensively used test for intelligence and —whosel thinking were influence. AMERICAN APPLIED PSYCHOLOGY1. James Mckeen Cattel— father of mental testing.2. Thorndike— father of educational measurement.3. Wechsler— adult intelligent and deviation quotient4. Raymond Cattel— advance statistical technique, culture free or culture fair-intelligence test.5. Safran— culture reduce intelligence test
9. DEVELOPMENT OF ACHIEVEMENT TEST1. Horace Man—written and oral examination2. Fisher— first objective test3. J.M. Rice— comparative test4. Stone— first standardize test in arithmetic5. Call— objective test which is extensively used today6. Taylor— evaluative test7. Gates— basic reading test for grade 3-8DEVELOPMNET OF CHARACTER AND PERSONALITY MEASUREMENT1. Fernald— first to measure character by test2. Voelker— some actual situations for testing character3. Symonds— scientific study of personality4. Rorschach—personality test using ink blots
10. PURPOSE OF EDUCATIONAL TESTING AND MEASUREMENT Instructional ---test is given to assess students' progress in a subject Curricular ---given when decision are to be made about school curricula. Selection ---test is given to determine students ability/suitability to enter school.(college entrance test)Placement --- test is given to group students (below average, geniuses, morons or homogeneous or heterogeneous)Personal --- assist individual to make wise decision for themselves (personality inventory test, aptitude test and its all standardize.)
11. ASSESSMENT PRINCIPLES 1. ADDRESS LEARNING TARGETS/CURRICULAR GOALS *cognitive*psychomotor*affective 2. PROVIDE EFFICIENT FEEDBACK ON INSTRUCTION *satisfactory (proceed to next) *unsatisfactory (reteach) 3. USE A VARIETY OF ASSESSMENT PROCEDURES *knowledge * skills *attitudes 4. ENSURE THAT ASSESSMENTS ARE VALID, RELIABLE, FAIR AND USABLE, Valid- reflects PURPOSE of the test Reliability- yields CONSISTENCE on the results Fair- fee from BIASES Usability- PRACTICABILITY, coverage, convenience, economical 5. KEEP

RECORD OF ASSESSMENT *allow student to documents their performance (portfolio) 6. INTERPRET/COMMUNICATE THE RESULTS OF ASSESSMENT MEANINGFULLY *test with correct meaning, student can make correct decision, failing scores can motivate, passing can inspire

12. CHARACTERISTICS OF MODERN ASSESSMENT •Objectives •Reliable •Multidimensional in structure • Measures knowledge, skills and values •Value-laden
13. FOUR KINDS OF STANDARD (ZAIS, 1976)ABSOLUTE MAXIMUM STANDARD (AMXS) -- attain by few students(95/100)ABSOLUTE MINIMUM STANDARD(AMNS) -- attain by majority to 75/100 guarantee promotion)RELATIVE STANDARD (RS) --- competence compared to other MULTIPLE STANDARD --- level of performance or competencies, RANK. Combination of AmxS and AmnS and RSEmploy to determine growth and pattern
14. INSTRUCTIONAL OBJECTIVES/LEARNINGOUTCOMES-- COMPOSED OF TWO ESSENTIALS 1.Learning outcomes -- ends results of instruction 2.Learning activities -- means to an ends Component: *behavioral- observables(include blooms taxonomy) and non-observables (know, appreciate, understand, value: to develop which are so general) •Content– specify the topic expected to learn. •* condition– words such, given, using, etc... •*criterion level– examples 75/100 acceptable level of performance
15. TAXONOMY OF INSTRUCTIONAL OBJECTIVES COGNITIVE DOMAIN (KCAPANSE) 1. Evaluation 1. Recall of facts 2. Understanding 2. Synthesis 3. Use previously learned information in 3. Analysis new setting 4. Breaking down facts 4. Application of a whole 5. Combining or putting 5. Comprehension facts to gather to create new scheme 6. knowledge 6. Making judgments.
16. REVISED TAXONOMY OF EDUCATIONAL OBJECTIVE Knowledge (product) Process 1. Remembering—retrieve info.. 1. Factual knowledge memory 2. Understanding– constructs 2. Conceptual summarizing, explaining knowledge 3. Apply-- executions, implementing 3. Procedural knowledge 4. Analyze– differentiate, integrate, attributing, 4. Meta-cognitive deconstructing, dissecting 5. Evaluate– judgments, checking, testing, and critiquing 5. Comprehension 6. Create – hypothesizing, designing, planning, 6. Knowledge constructing

17. AFFECTIVE DOMAIN
 1. Receiving ----willingness to receive (your teacher ask you to come in school early) then you accept
 2. Responding ----active participation, response in satisfaction (you comply to your teacher.)
 3. Valuing ----acceptance of value: preference and commitment (you believe the value of getting ahead of time)
 4. Organization ---- conceptualization of value and organizing of a value system
 5. Characterization ----- value system internalize value. Then the value became the character of the person.
18. PSYCHOMOTOR DOMAIN (RBPSN)
 •Reflex •----walking ,•*basic running, jumping•*perceptual•* skilled •----- dance, sports etc...•*non-discursive (non-verbal) • ----- gestures, sign language, pantomimes, body languages
19. CLASSIFICATION OF TEST A. According to purpose,/Uses
 1.Instructional
 2. Guidance
 3.administrative
 B. Format
 1. Standardize
 2. Teacher made test
 C. language mode.
 1. Verbal test
 2. Non-verbal test
20. GENERAL CLASSIFICATION OF TEST ITEMS
 1. Selection-typed item (student requires to select answer) --- multiple choices, true or false, matching type
 1. Supply-typed items (students requires to supply answer) ----- essay, short answer
21. KINDS OF TEST
 1. Intelligence test–
 2. Personality test
 3. Aptitude test
 4. Achievement test
 5. Prognostic test
 6. Performance test
 7. Diagnostic test
 8. Preference test
 9. Accomplishment test
 10. Scale test
 11. Speed test
 12. Power test
 13. Objective test
 14. Teacher-made test
 15. Formative test
 16. Summative test
 17. Placement test
 18. Standardize test
 19. Nor-reference test
 20. Criterion-reference test
22. INTELLIGENCE TEST MEASURES I.Q.
 1. Stanford-Binet Int. test— measures human ability, personality, characteristics, attitudes and interest.
 2. Wechsler-Adult Intelligence scale—(WAIS) verbal and non-verbal intelligence for adult
 3. Wechsler Intelligence Scale for children (WISC)– used for 5-15 years
 4. Safran Cultures-Reduce Intelligence Test— 36 items for children
 5. Culture free or Cultured Fair Intelligence Test– non-verbal intelligence test with two forms A and B consists of 50 items.
 6. Sequin Form-Board Test— sensory-motor skill of mental defectiveness
23. PERSONALITY TEST MEASURES INDIVIDUAL INTEREST COMPARE TO OTHERS
 1. Rorschach test with series of 10 ink blots
 2. Sixteen Personality Factor score-able test for getting the insight of person's personality

24. APTITUDE TEST-- PREDICTS WHERE THE STUDENT WILL LIKELY SUCCEED.
 1. Differential Aptitude test (DAT) -measures which field does student excel
 2. House-tree Person Test(HTP) – determines which of the tree an individual test
25. ACHIEVEMENT TEST--MEASURES WHAT HAS BEEN LEARNED BY THE STUDENT OF THE SUBJECT TAUGHT IN SCHOOL
26. PROGNOSTIC TEST-- Predict HOW WELL A STUDENT IS LIKELY TO DO IN A CERTAIN SCHOOL SUBJECT OR TASK.
 1. IOWA Placement Examination --foretells which of the subjects in the curriculum an examinee is doing good
27. PERFORMANCE TEST -- MAKES USE OF MANIPULATIVE MATERIALS WHICH INVOLVE MINIMUM VERBAL INSTRUCTIONS
 1. Field demonstration/internships/ etc.....
 2. Army beta
 3. Koh Block design
28. DIAGNOSTIC TEST --IDENTIFIES THE WEAKNESS AND STRENGTH OF AN INDIVIDUAL --it is usually given before instructions
29. PREFERENCE TEST --MEASURES VOCATIONAL INTEREST OR AESTHETIC JUDGMENT
 1. Kuder-reference record
30. ACCOMPLISHMENT TEST -- MEASURE OF ACHIEVEMENT USUALLY FOR INDIVIDUAL SUBJECT IN THE CURRICULUM
31. SCALE TEST -- SERIES OF ITEMS ARRANGED IN ORDER OF DIFFICULTY
 1. Binet-Simon Scale test this test was constructed from easiest to most difficult.
32. SPEED TEST -- MEASURES SPEED AND ACCURACY OF THE EXAMINEE WITHIN THE TIME LIMIT IMPOSED
33. POWER TEST --SERIES OF ITEM GRADED IN DIFFICULTY
34. OBJECTIVE TEST -- TEST WITH DEFINITE ANSWER
 1. Multiple choices
 2. Completion
 3. Enumeration
 4. Matching type
 5. True or false
35. TEACHER-MADE TEST -- CONSTRUCTED BY A TEACHER Can you give example.....!!!!!!
36. FORMATIVE TEST --USED TO MONITOR STUDENT ATTAINMENT OF THE INSTRUCTIONAL OBJECTIVE. ---test usually given after instruction.....note do not be confused between formative and summative (see next slide).....

37. **SUMMATIVE** -- DONE AT THE CONCLUSION OF INSTRUCTION AND MEASURES THE EXTENT TO WHICH STUDENTS HAVE ATTAINED THE DESIRED OUTCOMES It is usually given or taken monthly....so what is the difference between formative and summative???
- will you please save it.....it is very useful. One more thing to remember this concept is when you pay your tuition fee and do a hair cut for boys they will go to take a summative test..... Isn't it?
38. **PLACEMENT TEST** -- TEST USED TO DETERMINE THE GRADE OR YEAR LEVEL THE PUPIL OR STUDENT SHOULD BE ENROLLED Grade one should take this examination.....
39. **STANDARDIZED TEST** -- ARE VALID, RELIABLE AND OBJECTIVE, IT WAS MADE BY EXPERT Example: L.E.T. now it's your turn to give at least 5!!!!!! Think of an expert made and valid test.....
40. **NORM-REFERENCE TEST** -- IS A TEST THAT IS SCORED ON THE BASIS OF STANDARD LEVEL OF ACCOMPLISHMENT BY THE WHOLE GROUP TAKING THE TEST It is all test taken national, regional, division Most student and teachers ask the difference between norm and criterion..... You should be aware of those..
41. **CRITERION-REFERENCE TEST** -- A MEASURING DEVICE WITH A PREDETERMINED LEVEL OF SUCCESS OR STANDARD ON THE PART OF THE TESTTAKERS It was taken periodically
42. **SURVEY TEST** -- SERVES A BROAD RANGE OF TEST Commonly used in thesis,,
43. **MASTERY TEST** -- SCORES SPECIFIC LEARNING OBJECTIVES
44. **SUBJECTIVE** -- OPPOSITE OF OBJECTIVE It is scored affected by bias...and it is common in the form of essay... Note: essay test can be objective...
45. **VERBAL TEST** -- TEST USING WORDS Dictation, puzzles

46. NON-VERBAL TEST -- TEST USE PICTURES OR SYMBOLS

47. STEPS IN DEVELOPMENT AND VALIDATION OF TEST (OTFWE) PHASE I 1.

Determining the objectives 2. Preparing the TOS 3. Selecting the appropriate test format 4. Writing the test 5. Editing the test item

48. PHASE II Test construction stage PHASE III 1. Administering the first try out then do ITEM

ANALYSIS 2. Administering the second tryout do ITEM ANALYSIS 3. Preparing the final form of the test then establish validity PHASE IV Evaluation stage Establishing test reliability Interpreting test scores

49. TABLE OF SPECIFICATION (TOS) -Blue print of the test -Represent learning outcomes to

be tested, percentage of items, item placement, type of test, number of item, and number of recitation. -BEHAVIOR- -No. of days -BEHAVIOR- CONTENT --- no. of teacher CONTENT ---learning outcome taught particular ---is equals to the to be tested topic number of recitation over total no. of recitation day for the-Percentage of item -Item placement whole quarter times--- no. of item per --- reference of total no. of item's objective over total particular entry item no. of item

50. TYPES OF TOS ONE-WAY GRID Obj... No. of No. of items percentage Item recitation

placement TWO-WAY GRID Obj. K Co Ap An Sy Ev No. of No. of Perc. Item Type total rec. item placem ent 1.2.Total no of item perc.

51. WRITING TEST ITEM'S Click on the underline words to see links 1. Multiple Choice 2.

True or False 3. Matching Type 4. Restricted response Item or completion Test 5. Essay

52. CHARACTERISTICS OF A GOOD TEST a. Valid b. Reliability a. It is presented in a new

slide for fast and to refresh both mind and computer.....!!!!!!!

S.sr test construction admin and scoring Presentation Transcript

- 1. Test Construction, Administration, and Scoring Victor P. Bactol, Jr.
- 2. What to Discuss • Test Construction • Test Administration • Test Scoring
- 3. TEST Construction • Determine what is to be measured • Create instruments that will provide the best measure • Planning a Test • Preparing a Test • Assembling a Test
- 4. What is to be measured? • The amount of efforts in the construction of educational or psychological test varies with consideration of type and purpose of the test. • Classroom teachers spent little time for the preparation for essay or short-answer test. • Complex procedures followed by professional test designers are unfamiliar to the majority of teachers.
- 5. • Whatever kind of test or goals, users have to do some degree of planning. • Create instruments that will provide the best measure
- 6. Planning a Test • Questions for Test Planners: 1.What are the topics and materials on which the students are to be tested? 2. What kind of questions should be constructed? 3. What item and test formats or layouts should be used?
- 7. 4. When, where, and how should the test be given? 5. How should the completed test papers be scored and evaluated? • Note: The first 3 questions pertain to test design and construction, the 4th question to test administration, and the 5th question to test scoring.
- 8. Taxonomies of Educational Objectives • The preparation of a test to measure specific instructional objectives is most effective when the behaviors to be assessed are clearly defined at the outset. • Recently, there are formal standard systems for classifying the cognitive and affective objectives of educational instruction.
- 9. 1. Cognitive Domain: • Bloom and Krathwohl's (1956) – The cognitive domain are listed in order from least to most complex. These six categories are not exclusive but rather progressively inclusive. • Educational Testing Service (1965) • Gerlach and Sullivan (1967) • Ebel (1979)
- 10. Categories of the Taxonomy of Educational Objectives: Cognitive Domain • Knowledge involves the recall of specific facts. Sample verbs in knowledge items are define, identify, list, and name. • understanding the meaning or purpose of something. Sample verbs in comprehension items are convert, explain, and summarize.

- 11. • Application involves the use of information and ideas in new situations. Sample verbs in application items are compute, determine, and solve. • Analysis is breaking down something to reveal its structure and the interrelationships among its parts. Sample verbs are analyze, differentiate, and relate. • Synthesis is combining various elements or parts into a structural whole. Sample verbs are design, devise, formulate, and plan. • Evaluation is making a judgment based on reasoning. Sample verbs are compare, critique, and judge.
- 12. • Other taxonomies may be followed • Note: Following any given outline of Cognitive Objectives should encourage the test preparer to go beyond simple cognitive or rote memory items and construct a number of test items to measure higher order educational objectives requiring some thought. •
- 13. 2. Affective Domain: • - Another important function of education is instilling certain attitudes, values, and other affective states in the learner. • - A completely satisfactory method of classifying the affective objectives of instruction does not exist, but proposals have been made. • Krathwohl, Bloom & Masia, 1964
- 14. - Taxonomies of instructional objectives in the psychomotor domain have also been proposed: • Simpson, 1966 • Harrow, 1972 – The six categories in Harrow’s Taxonomy of Psychomotor Domain: Reflex Movements, Basic- Fundamental Movements, Perceptual Abilities, Physical Abilities, Skilled Movements, and Nondiscursive communication. •
- 15. 3. Table of Specifications: • Constructing a table of specification is helpful in planning a test. • Once a set of objectives for a course of study has been decided on and a topical outline prepared, test items can be constructed to measure the extent to which students have attained the objectives listed for each topic.
- 16. • It is referred to in deciding what varieties of items, and how many of each, are appropriate. • Many practical considerations – cost, time, available for administration, item arrangement, testing conditions, and the like – must also be taken into account in planning a test. • It serves as a guide in constructing items to assess (or predict in the case of an aptitude test) certain objectives.
- 17. Preparing the Test Items • Construction of various items can be had once a table of specifications or other fairly detailed outline of the test has been prepared. • It is recommended

that, on objective tests, about 20% more items than are actually needed be written so that an adequate number of good items will be available for the final version of the test.

- 18. • Various methods of classifying test items according to their format or the form of response required have been suggested: Supply versus selection, recall versus recognition, and constructed response versus identification are all ways of differentiating between those items. • Another popular method of classifying items is essay versus objective. All essay items are of supply type. Objective items however, may be either the supply or selection type. Examples are provided below:
- 19. I. Essay Items Direction: Write a half page answer to each item. 1. Contrast the advantages and disadvantages of essay and objective test items. 2. Explain the reasons for performing an item analysis of a classroom test.
- 20. II. Objective Items A. Short-answer Directions: Write the appropriate word/s in each blank 1. The only thing that is objective about an objective test is the _____. 2. What is the first formal step in constructing a test to predict degree of success on a particular job?

- 21. B. True-false Directions: Circle T if the statement is true; circle F if it is false. T F 1. The most comprehensive test classification system is that of The Mental Measurement Yearbooks. T F 2. The social-desirability responses set is the tendency to rate an examination high on one trait simply because he/she is rated high on another trait.
- 22. C. Matching Directions: Write the letter corresponding to the correct name in the appropriate marginal dash. _____ 1. group intelligence test A. Binet _____ 2. individual intelligence test B. Darwin _____ 3. interest inventory C. Galton _____ 4. personality inventory D. Otis _____ 5. product-moment correlation E. Pearson _____ 6. sensorimotor tests F. Rorschach G. Spearman H. Strong I. Woodworth
- 23. D. Multiple-choice Direction: Write the letter of the correct option in the marginal dash opposite the item. ____ 1. Qualifying words such as never, sometimes, and always, which reveal the answer to an examinee who has no information on the subject of the item, are called A. glittering generalities B. interlocking adverbs C. response sets D. specific determiners

- 24. ____ 2. Jimmy, who is 8 years, 4 months old, obtains a mental age score of 9 years, 5 months. What is his ratio IQ on the test? A. 88 B. 90 C. 113 D. 120
- 25. 1.Characteristics of Essay Items • they can measure the ability to organize, relate, and communicate – behaviors not easily so easily accessed by objective items • susceptibility to bluffing by verbally fluent or facile but uninformed examinees. • the scoring of essay tests is rather subjective and time consuming.
- 26. Rule: An essay item should not be used when it is possible to measure the same thing with an objective item. • If essay questions are to be asked, the item writer should try to make the questions objective as possible. • This can be done by (1) defining the task and wording the items clearly (e.g. asking examinees to “contrast” and “explain” rather than “discuss”). (2) using a small number of items , all of which should be attempted by all examinees and (3) structuring the items in such a way that subject-matter experts will agree that one answer is better than another.
- 27. Other Types of Objective Items - there are other types of objective items other than traditional four (short- answer, true-false, matching, and multiple- choice), but these four are certainly the most popular. • Advantages: - they can be easily and objectively scored - because less time is needed to answer an item, they permit a wider sampling of material than essay tests. •
- 28. • Rule: - in preparing objective tests, care should be taken to make the items clear, precise, grammatically correct, and written in language suitable to the reading level of the group for whom the test is intended. - all information and qualifications needed to select a reasonable answer should be included, but nonfunctional or stereotyped words should be avoided.
- 29. - lifting statements verbatim from the textbooks or other sources should be avoided; this practice puts a premium on rote memory. - exclude irrelevant clues to the correct answers to avoid interrelated or interlocking items. Interrelated Items are those on which the wording of one item gives clue to the answer to another item. Interlocking Items are those in which knowing the correct answer to one item will get the other item right.
- 30. a. Short-answer Items – a supply type item: examinees are required to complete or fill in the blanks of an incomplete statement with the correct word/s or phrase or to give a brief answer to a question . - fall somewhere between essay and recognition - easiest items - requires examinees to supply than recognize - useful in assessing knowledge of terminology but - they serious

limitations - cannot measure more complex instructional objectives - because they have more than one correct answers, they cannot be scored objectively.

- 31. • Guidelines: - Use question rather than incomplete statements - If an incomplete statement is used, word it so that the blank comes at the end. - Avoid multiple blanks in the same item, especially if they make the meaning of the task unclear.
- 32. b. True-False Items – One of the simplest types of items to construct. - can be written and read quickly; - permit a broad sampling of content. - but they often deal with trivia or - constructed by lifting statement verbatim from the textbooks. - more complex objectives are difficult to measure adequately with this kind of items. - test result may be greatly affected by the examinee's tendencies to guess and to agree or disagree when in doubt. - the meaning of the score may be questionable.
- 33. - on the average the examinees will get 50% of the items correct by simply guessing. - scores may be inflated even more when items contain specific determiners – words such as all, always, never, and only which indicate that the statement is probably false. - or often, sometimes, and usually which indicate that the statement is probably true.
- 34. - In addition to avoid specific determiners; the following precautions are advisable in writing true- false items: 1) Make the statement relatively short and unqualifiedly true or false. 2) Avoid negatively stated items, especially those containing double negatives, as well as ambiguous and tricky items. 3) On opinion question, cite the source of authority. 4) Make the true and false statements about the same length, and make the number of the true statements approximately equal to the number of false statement.
- 35. c. Matching Item – In a sense, both true-false and multiple- choice items are varieties of matching items. On both types of items, a set of response options is to be matched to a set of stimulus options. - the following are the guidelines to follow: 1) Place the stimulus and the response options in a clear, logical order, with the response options on the right. 2) List between six and fifteen options, including two or three more response options than stimulus options 3) Clearly specify the basis for matching 4) Keep the entire item on a single page.
- 36. d. Rearrangement items – are type of matching items on which the examinee is required to sort a group of options into a fixed number of predetermined categories. The ranking items in

which the options are arranged in rank order from first to last is a special type of rearrangement item.

- 37. e. Multiple-choice items – the most versatile form of objective item. - it can be used to measure both simple and complex learning objectives, and scores on multiple-choice items are less affected by guessing and other response sets than are scores on other types of objective test items. - furthermore, useful diagnostic information may be obtained from an analysis of the incorrect options (distracters) selected by an examinee. - one shortcoming of multiple –choice items is that good ones are difficult to construct, especially items on which all options are equally attractive to examinees who do not know the correct answer.
- 38. - the following suggestions prove helpful in constructing the stems and options of a high-quality multiple choice items: 1) Either a question or an incomplete statement may be used as the stem for a multiple-choice item, but the stem should ask the question or state the problem clearly. 2) As much of the item as possible should be placed in the stem, since it is inefficient to repeat the same words in every option. 3) Four or five options are typical on multiple-choice items, but good items having only two or three options can also be written.
- 39. 4) All options should be approximately the same length and grammatically correct in relation to the stem. 5) If the options have a natural order, such as dates or ages, it is advisable to arrange them in that way; otherwise, the options should be arranged in random order or alphabetized (if alphabetizing does not give clues to the correct answer). 6) All options should be plausible to examinees who do not know the correct answer, but only one option should be correct. Popular misconceptions or statements that are only partially correct make good distracters. •
- 40. Writing Complex Items • Test constructors usually have more difficulty writing items to measure understanding and thinking process than straightforward knowledge of the test material. • there are, however, various ways of constructing objective test items to measure the more complex objectives of instructions. • Example: including two or more propositions in the statement or stem of a multiple –choice or true-false item can increase the difficulty level of the item – increasing the complexity of Multiple-choice.
- 41. Multiple Premises: Given that Mary’s raw score on a test is 60, the test mean is 59, and the standard deviation 2, what is Mary’s score? (a) - 2.00 (c) .50 (b) - .50 (d) 2,00

- 42. Classification: Jean Piaget is best characterized as a(n) _____ psychologist. (a) developmental (c) psychometric (b) industrial (d) social
- 43. Oddity: Which of the following names does not belong with the others? (a) Adler (c) Jung (b) Freud (d) Rogers
- 44. Multiple True-false: Is it true that (1) Alfred Binet was the father of intelligence testing, and (2) his first intelligence test was published in 1916? (a) both 1 and 2 (c) not 1 but 2 (b) 1 but not 2 (d) neither 1 nor 2
- 45. Relations and Correlates: Mean is to standard deviation as median is to (a) average deviation (c) semi-interquartile range (b) inclusive range (d) variance
- 46. If . . . Then: If the true variance of a test increases but the error variance remains constant, then what will be the effect? (a) test reliability will increase (b) test reliability will decrease (c) observed test variance will decrease (d) neither test reliability nor observed variance will be changed
- 47. Assembling a Test - Review and edit the test items by another knowledgeable person (friend, or associate) to spot errors and valuable suggestions for improving items. - final decisions concerning several matters must be made: 1) Is the length of the test appropriate for the time limits? 2) How should the items be grouped or arranged on the pages of the test booklet?
- 48. 3) Are answers to be marked in the test booklet, or is a special answer sheet to be used? 4) How will the test booklet and answer sheet be produced ? 5) what should be included in the directions?
- 49. The Length - The decision of how many items to include on a test depends on the time limits and the grade and reading level of the examinees. - For secondary school level and beyond, a good rule is to allow 1 minute per item on multiple- choice and short –answer tests and 1 minute per two items on true-false tests.
- 50. - a 50-item multiple-choice or short answer test and a 100-item true-false test are satisfactory from a typical class period. - About five or six half-page essay questions can be answered in this same period. - with these length lengths , 80 percent or more of the students in a typical secondary school or college-level class can be expected to finish the test. - when testing elementary school pupils, these suggestions concerning test length will need to be revised.

- 51. Arrangement of Items - Multiple-choice or true-false items should be arranged in the test booklet so that the answers follow no set pattern. - Generally it safer to randomize the order of alternatives. - Placing short-answers items in groups of five or so reduces errors in taking and scoring the tests. - On matching or rearrangement items, all the options should appear in the on the same page .
- 52. - If answers to short-answer and essay items are to be written in the test booklet, sufficient space must be provided. - grouping items of the same type and the same topic will facilitate the examinee's task. - it is reasonable to suppose that test scores would be higher if subsets of items are arranged in order from easiest to most difficult. - arranging items in groups according to type may make test preparation, administration, and scoring easier.
- 53. Answer Sheets - Test Booklet. This will lead to fewer error s by examinees in finding the appropriate space for a given item. - Marginal space to the left of the question will also help. - Separate answer sheet which easier to score - can be used in upper elementary and beyond - If the answer sheets are to be scored by hand, the teacher can reproduce the quantity.
- 54. - answer sheet for a fifty-item multiple-choice test might have a format like this: 1. A B C D E 26. A B C D E 27. A B C D E 25. A B C D E 50. A B C D E - Examinees would be directed to check, blacken, or circle the letter corresponding to the correct answers to each item.
- 55. Reproducing a Test - Reproducing the copies of materials may be done through the use of machine/s available: Mimeograph, Ditto, Photocopy, Offset machine, Rizo, etc.
- 56. Test Directions - tell the examinees what they are supposed to do and how long they have in which to do it. - should indicate in relatively simple language, the purpose of the test, the time limits, the method of recording answers and the way the test is to be scored. - they should also indicate whether examinees should hues at answers when in doubt. - the following is an example of the general directions for a classroom achievement test:
- 57. Write your name in the upper right-hand corner sheet; do not write anything in the test booklet. The purpose of this test is to determine your knowledge and understanding of test preparation, administration, and item analysis. There are 50 questions, and you will be given 50 minutes in which to complete the test. Indicate your answers on the separate answer sheet by marking a check through the letter corresponding to the correct answer to the item; do not write

your answer in the test booklet. Your score on the test will be the total number of items that you answer correctly. Therefore, you should make an informed guess when in doubt about the correct answer. Do not omit any items. If you have questions during the test, raise your hand. If you finish before time is called, sit quietly until everyone has completed the test. •

- 58. Oral Testing - an evaluation situation in which responses to questions are given orally by examinees, the questions themselves being presented orally, in writing, or both. - students do not like oral tests and feel they are unfair measures of knowledge and understanding. But,
- 59. - teachers of speech, dramatics, English, and foreign languages often decry the current inattention to the assessment of spoken language skills and feel that the consequence of this neglect is a multitude of graduates who cannot speak correctly, comprehensively, and comfortably. - but there are teachers in languages that point out to the inefficiency and subjectivity of oral testing. (Platt, 1961; Crowl & McGinitic, 1974). - since the early part of this century, oral achievement tests have tended to be perceived as inefficient and lacking in psychometric rigor.
- 60. Advantages: - interactive social situation provided by oral examinations, permitting the evaluation of personal qualities such as appearance, style, and manner of speaking. - the face-to-face situation also makes cheating, and perhaps bluffing less likely on oral tests. - they provide practice in oral communication and social interaction. - require responses at higher intellectual level than written tests. - oral examiners may be better able to follow the thought processes of examinees and determine the boundaries of their knowledge and understanding of the subject more readily.
- 61. Oral versus Written Examination - the two types of evaluation measure different aspects of achievement. - in general, knowledge of specific facts can be assessed more quickly with objective written tests, so oral examinations should not contain large number these kinds of questions. - with essay test s, oral testing is more appropriate when the questions call for extended responses. - main source of error in oral resting – the examiners themselves. They should have a thorough knowledge of the subject and be keenly aware of the kinds of oral responses desired.
- 62. The Rules • Use test to measure appropriate outcomes • Make test comprehensive; exclude trivial • Avoid recall or recognition • Choose best test item for objective • Avoid trick questions •

Use large number of items • Do not use book statements • Avoid intra-test references • Make items independent • Avoid obvious questions • Avoid specific determiners • Use simplest possible wording • Strive for clarity; eschew ambiguity • Work toward simple data recording • Leave sufficient space • Keep responses on right • Use separate answer sheet when possible • Avoid answer patterns • Keep questions on single page • Number consecutively • Increase objectivity by underlining • Indicate corrections for guessing • Make clear and complete directions • Make a proper heading

- 63. TEST Administration • How can test administration affect validity and/or reliability? • Distractions • Cheating • Instructions • Supervision
- 64. Administering the Test Note: - Certain examiner and situational variables also influences test performance. - The degree of preparedness, the personality, and the behavior of an examiner during the test are examiner variables affecting examinee performance. - Other situational variables affecting test scores are the place, time, and environmental conditions.
- 65. 1. Examiner's Duties before the Test a) Announcing the test b) Becoming familiar with the test c) Ensuring satisfactory testing conditions d) Minimizing Cheating
- 66. Announcing the Test - examinees should be informed beforehand: • when and where the test will be administered, • with what content it will deal • what sort of test (objective, essay, oral) - Students deserve a chance to prepare themselves for tests, intellectually, emotionally, and physically. - for this reason, "pop quizzes" and other unannounced tests are usually inadvisable.
- 67. Becoming Familiar with the Test - If the test constructor and the examiner are the same person, there is no question of the latter's familiarity with the test material and administration procedure. - But because administrator of standardized test is rarely the same person who constructed the test, the latter will need to study the accompanying manual carefully before attempting to give the test.
- 68. - Direction of the administration must be understood - Content of the test should be familiar to the examiner. - To attain familiarity, it is advised that the examiner should also take the test before attempting to give the test. - Directions and other procedures should be reviewed before the administration of the test.

- 69. Ensuring Satisfactory Testing Conditions - To make certain that seating, lighting, ventilation, temperature, noise level and other physical conditions are appropriate. - Special provisions may be made for examinees who have physical handicaps or are physically different from most other examinees. •
- 70. Minimizing Cheating - Comfortable seating that minimizes cheating should be arranged. - although preferred, it is not always possible to seat examinees one seat apart in such a way that cheating is impossible. - Preparing multiple forms (different items or different item arrangement) - Several proctors should be employed whenever a large group of people are tested. - Other procedures designed to guard against cheating must be taken quite seriously in the administration of “secure” standardized tests such as the Scholastic Aptitude Test and the Graduate Record Examinations. •
- 71. 2. Examiner’s Responsibilities during the Test a) Following Test Directions b) Establishing Rapport c) Being Prepared for Special Problems
- 72. Following the Test Directions - Examiner is asked to follow the directions for administration carefully even when further explanation to examinees might clarify their task. - Departures from the standard directions may present a different task to examiners than the test designers had in mind.
- 73. Establishing Rapport - Examiner must have a behavior that tends to create a condition of rapport, a relationship between examiner and examinees that encourage the latter to do their best.
- 74. Being Prepared for Special Problem - A test situation creates a certain amount of tension in almost everyone, and occasionally an examinee may become quite anxious, etc. - The examiner must be alert, flexible, warm, and objective, as well as familiar with the test material. - Certain measures coupled with sensitivity and patience on the part of the examiner can provide better opportunity for the handicapped individuals and those with problems to demonstrate their capabilities.
- 75. TEST Scoring Note: Teacher may wish to determine in advance separate scores on various parts as well as composite score on the test as a whole. - differential numerical weights may be assigned to different responses.

- 76. - decisions must also be made on such a matters as whether to subtract a correction for guessing and whether to report the results in raw score form or to convert them in some way. - For standardized tests classroom teacher does not have to make all these decisions; the answer sheets can be scored by machine.
- 77. Scoring the Test • Make an answer key • Make sure scoring marks are visible • Write student's score on their test – if it will be returned to them
- 78. 1. Scoring Essay Tests - By structuring the task clearly so that interpretation of a question will not vary widely from examinee to examinee. - Can be based on the quality of the answer. - an attempt can be made to structure and objectify the scoring of essay items so that a person's score depends less on non content, impressionistic factors and more on the understanding and ability demonstrated. - Scoring on the basis of penmanship rather than on quality of answers has the so-called (halo effect) among the errors affecting scores on essay tests.
- 79. Recommendations (so that scores will be as objective and reliable as possible) - The tester must decide whether to score the question as a whole or - Assign separate weights to different components. - Whole (global) scoring is common, but it is perhaps more meaningful to use analysis scoring procedure in which points are given for each item of information or skill included to an answer.
- 80. - It is advisable for the examiner to write out an ideal answer to the question beforehand. - The names of examinees must be blocked out, if possible, before the test papers are scored. - Score all examinees' answers to one question before going to the next question. - After a second person rescores the papers, make the final score the average of the number of points assigned by the two scorers. •
- 81. 2. Scoring Objective Test - A unique advantage of objective test is the efficiency and objectivity with which they can be scored. - A clerk can score objective tests quickly and accurately with a scoring stencil or machine. Thus the test papers can be returned to the examinees soon. - A scoring stencil for use with a special answer sheet can be prepared from a blank sheet of paper or cardboard by punching out spaces where the correct answers should be. •
- 82. a) Scoring Weight for Multiple Machine and True- False Test - It is generally agreed that on tests of twenty or more items, simply assigning a score of 1 to each correct response and 0 to each incorrect response is just as satisfactory as using different scoring weights. - Thus a

multiple-choice or true-false test of fifty items, the range of possible score would be from 0 to 50.

- 83. b) Scoring Other objective Items - As with true or false and multiple-choice items, short-answer and matching items may be scored by assigning 1 point to each correct response and 0 point to each incorrect or omitted response.
- 84. c) Correction for Guessing - Examinees do guess on objective tests and their chances of improving their scores in this way, especially on items having few options, can be quite high. Note: chances of that the correct option will be selected by guessing are $100/k$ out of 100, where k is the number of option per item. Thus: – In true-false items – 50 out of 100 ($k=2$ for true-false items) – On a four option multiple-choice item the chances are only 25 out of 100 of guessing the correct answer. – Therefore obviously guessing will have more serious effect on true-false items rather than on multiple-choice items.
- 85. - An attempt to correct for the effects of guessing, on certain standardized tests a portion of the number of wrong answers is subtracted from the number of right answers. - The most popular formula when correcting for guessing is considered appropriate is: $S = R - W/k$ – S – corrected score R – number of items that the examinee gets right W – the number of items wrong
- 86. Note: Professional testers generally agree that so-called correction-for-guessing formula do not really correct for guessing and that they typically have little effect on the order of scores.
- 87. d) Converted Scores - Although it may be worthwhile to alter raw scores on objective tests by weighting or correcting formulas, the scores are often changed to make them more meaningful. - The interpretation of test scores is facilitated by transforming them to standard scores, percentiles, or other converted scores.
- 88. e) Scoring Oral Tests - There are greater possibilities of error in scoring responses to oral test. - To improve the objectivity of scoring oral tests – recommendations to be followed: - careful attention to the design of question - construction of model answer - use of multiple raters or scores - training examiners to avoid showing favoritism and other rate biases - if time allotted to scoring is not critical the accuracy with which oral tests are scored can be improved by recording the examinees' responses.

- 89. Test Results • Raw score, Percentage, Percentile • Range of scores – difference between highest and lowest scores. – $\text{Range} = \text{SH} - \text{SL} + 1$ – What does this tell us? • Measures of Central Tendency – Mean – Median – Mode
- 90. Analyzing Test Results • Step 1: Arrange scores highest-lowest • Raw scores, %, %ile • Range of scores • Measures of central tendency • Sort tests into ascending order by score; divide into quarters – What info can we get from this?
- 91. Analyzing Test Results Step 2: Analyze each item – Item analysis • Difficulty Factor = C/N – Target 0.40 to 0.90 – Range: 0 (no correct answers) to 1 (all correct answers) • Discrimination Index = $(CH - CL)/nH$ – Target 0.30 to 0.70 – Positive for correct answer, negative for incorrect – Range: -1.0 to +1.0 • Kuder-Richardson 20 – Target 0.80 to 0.85 What info can we get from these?
- 92. Analyzing Test Results • Step 3: Act appropriately on the analysis – Eliminate items and adjust scores – Class discussions/re-teaching – Improve assessment instrument