ASSESSMENT OF LEARNING

BLEPT COMPETENCIES: ***(from NCBTS)***

1. Apply principles in constructing traditional and alternative/authentic forms of high quality assessment
2. Interpret assessment result
3. Utilized processed data and assessment results in reporting learners' performance to improve teaching and learning
4. Demonstrate skills in the use of techniques and tools in assessing affective learning
5. Assign students' marks/ratings

CONTENT UPDATE

BASIC CONCEPTS

Test

- An instrument designed to measure any quality, ability, skill or knowledge
- Serves as a vehicle used to observed traits

Measurement

- The assignment of numbers to certain characteristics
- A process of quantifying the degree to which someone/something possesses a given trait

Assessment

- Related series of measures used to determine a complex attribute of an individual
- A process of gathering and organizing data into an interpretable form to have basis for decision-making

Evaluation

- A systematic analysis of both qualitative and quantitative data in order to make sound judgment or decision
- Refers to the consideration of evidence in light of value standard

MODES OF ASSESSMENT

MODE	DESCRIPTIONS
Traditional	▪ The objective paper-and-pencil test which usually assesses low-level thinking ▪ *Tools:* Standardized tests, objective test ▪ *Advantages:* (1) scoring is objective, (2) administration is easy because students can take the test at the same time ▪ *Limitations:* (1) Preparation of instrument is time-consuming, (2)prone to cheating
Performance	▪ A mode of assessment that requires actual demonstration of skills or creation of products of learning ▪ *Tools:* Practical test, oral recitations, projects ▪ *Advantages:* (1) preparation of the instrument is relatively easy, (2) measures behaviors that cannot be deceived ▪ *Limitations:* (1) Scoring tends to be subjective without rubrics (2)Administration is time-consuming
Portfolio	▪ A process of gathering multiple indicators of student progress to support course goals in dynamic, ongoing and collaborative process ▪ *Tools:* Working portfolio, show portfolio, documentary portfolio ▪ *Advantages:* (1) measures students' growth and development, (2) intelligence-fair ▪ *Limitations:* (1) development is time consuming, (2) rating tends to be subjective without rubrics
Affective	▪ Used to assess affective targets ▪ Tools: Teacher observation, student self-report, peer rating ▪ Advantages: (1) assesses interpersonal skills, affective objectives and classroom environmental targets ▪ Limitations: (1) preparation of instrument is time-consuming, (2) observation tend to be distorted if not promptly recorded

TYPES OF TEST

PSYCHOLOGICAL
- Measures students' intelligence or mental ability in a large degree without reference to what the students has learned

EDUCATIONAL
- Aims to measure the result of instructions and learning

STANDARDIZED
- Constructed by an EXPERT
- Covers a broad range of content covered in a subject area

TEACHER-MADE/ INFORMAL
- Constructed by a CLASSROOM TEACHER
- Covers a NARROW range of content

SURVEY
- Covers BROAD range of objectives
- Measures general achievement in certain subjects

MASTERY
- Covers SPECIFIC objectives
- Measures fundamental skills and abilities

NON-VERBAL
- Examinees do NOT use words in attaching meaning to or in responding to test items

VERBAL
- Words are used by examinees in attaching meaning to or responding to test items

GROUP
- Typically, paper-and-pen test
- Same amount of time needed to gather information from one student

INDIVIDUAL
- Mostly given orally or requires actual demonstration of skill
- Chance to follow-up examinees' response

OBJECTIVE
- Personal judgment does not affect the scoring

SUBJECTIVE
- Affected by scorer's personal biases

POWER
- Measures students ability to answer more and more difficult items

SPEED
- Measures students' speed and accuracy in responding

SELECTIVE
- There are choices for the answer.
- Can be answered quickly.

SUPPLY
- NO choices for the answer
- May require longer time to answer

MAXIMUM PERFORMANCE
- Determines what individuals CAN DO when performing at their best

TYPICAL PERFORMANCE
- Determines what individual WILL DO under natural conditions

NORM-REFERENCED
- Emphasizes DISCRIMINATION among individuals in terms of level of learning
- Favors items of average difficulty

CRITERION-REFERENCED
- Emphasizes DESCRIPTION of what learning tasks individual can and cannot perform

PRINCIPLES OF HIGH QUALITY ASSESSMENT

1. Clarity of learning targets
   - Clear and appropriate learning targets include (1) what students know and can do and (2) the criteria for judging students' performance.
2. Appropriateness of Assessment methods
   - The method of assessment to be used should match the learning targets.
3. Validity
   - This refers to the degree to which score-based inference is appropriate, reasonable, and useful. A test is said to be valid if it measures what it supposed to measure.
4. Reliability
   - This refers to the degree of consistency when several items measure the same thing and stability when the same measures are given across time.
5. Fairness
   - Fair assessment is unbiased and provides students with opportunities to demonstrate what they have learned.
6. Positive consequence
   - The overall quality of assessment is enhanced when it has a positive effect on student motivation and study habits.
   - For teachers, high quality assessment leads to better information and decision-making about students. For
   - students, high quality assessment serves as a motivation and leads students to a better study habits.
7. Practicality and efficiency
   - Assessments should consider the teacher's familiarity with the method, the time required, the complexity of administration, the ease of scoring and interpretation, and cost.

FOUR TYPES OF EVALUATION PROCEDURES

TYPES	DESCRIPTIONS
1. Formative Evaluation	▪ Done **DURING** instruction ▪ Informs learners whether they have mastered a unit or not ▪ Purpose: Reinforce learning, diagnose strength and difficulties, handle errors, quality, control, forecast success ▪ Tools: Short quizzes, drills, seat works ▪ Generally, NOT grade
2. Summative Evaluation	▪ Done **AFTER** instruction ▪ Certifies mastery of intended learning outcomes ▪ Purpose: Assign grades, compare outcomes of different groups ▪ Tools: Unit/chapter test, quarter exams, final exams
3. Preliminary Evaluation	▪ Done **BEFORE** instruction ▪ Determines mastery of prerequisite skills ▪ Purpose: Determine readiness of students ▪ Tools: Placement tests, readiness test
4. Diagnostic Evaluation	▪ Done **BEFORE or/and DURING** instruction ▪ Judge students' level of performance ▪ Determines persistent difficulties and searches for their underlying causes ▪ Purpose: Find solutions and remedy to problems/difficulties ▪ Tools: Diagnostic test, oral recitations, board work, seat work

FOUR FRAMES OF REFERENCES FOR INTERPRETING ASSESSMENTS

REFERENCES	DESCRIPTIONS
Ability	Performance is interpreted in light of the student's maximum possible performance **Examples**: (1) This student can do better if given more time. (2) John did his best in this exam.
Growth	Performance is interpreted in comparison with student's prior performance **Examples**: (1) Joana has improved a lot. (2) Ken performed better than yesterday.
Norm	Interpretation is provided by comparing the student's performance with the performance of others **Examples**: (1) This student got the highest score in Physics.

	(2) Joana can encode the fastest in this class.
Criterion	Performance is interpreted by describing what the student can and cannot do. **Examples**: (1) In algebra, Eric can identify the missing term in a sequence. (2) Kweenie cannot dance the Gangnam Style.

-oOo-

CLASSIFICATION OF LEARNING OUTCOMES (Learning Taxonomies)

Instructional Objectives- statements that describe the abilities students should be able to display to demonstrate that important concepts and principles have been incorporated into their own structures of knowledge. (Raagas, 2010)

Enabling Objectives- prerequisite knowledge learners need to enable them to benefit from instruction.

A. Cognitive Domain (Bloom, et Al. 1956)

CATEGORIES	DESCRIPTIONS
Knowledge	▪ Recall of information ▪ **Some questions cues:** list, define, tell, describe, identify, show, label, recall, repeat, state, name, arrange, examine, tabulate, quote, collect **Example**: State all the categories of learning outcomes under cognitive domain.
Comprehension	▪ Grasp meaning of the material, understand information, and interpret facts ▪ **Some question cues:** summarize, describe, interpret, contrast, predict, associate, distinguish, estimate, differentiate, discuss, extend, classify, explain, identify **Example**: Differentiate measurement and assessment.
Application	▪ Use learned material in new and concrete situations, solve problems using required skills or knowledge ▪ **Some questions cues:** apply, demonstrate, calculate, complete, illustrate, show, solve, examine, modify, relate, change, classify, experiment, discover, choose ,dramatize **Example**: Using the Pythagorean theorem, solve the hypotenuse of at least five right triangles.
Analysis	▪ Break down material into its component parts so that its organizational structure may be understood, see patterns, recognition of hidden meanings ▪ **Some questions cues:** analyze, separate, order, explain, connect, classify, arrange, divide, compare, select, explain, infer, appraise, calculate, categorize, contrast, criticize **Example:** Compare and contrast formative and summative evaluation in terms of their description, tools and purpose.
Synthesis	▪ Put parts together to form a new whole, use old ideas to create new ones, generalize from given facts, predict, draw conclusion ▪ **Some questions cues:** combine, integrate, modify, rearrange, substitute, plan, create, design, invent, compose, formulate, prepare, generalize, rewrite, arrange, assemble **Example:** Create at least one instructional objective under each category of the cognitive domain.
Evaluation	▪ Judge the value of material for a given purpose, compare and discriminate between ideas, assess value of theories and presentations, make choices based on reasoned argument ▪ **Questions cues:** assess, decide, rank, grade, test , measure, recommend, convince, select, judge, explain, discriminate, support, conclude, compare, summarize, etc. **Examples:** (1) Using the guidelines of constructing objective tests, critique a given test item by identifying any error and giving a suggestion for improvement. (2) Classify a given a set of cognitive objectives based on their category.

B. Affective Domain (Krathwohl, et Al. 1964)

CATEGORIES	DESCRIPTIONS
Receiving	▪ Willingness to receive or to attend to particular phenomena or stimuli

	▪ **Illustrative verbs**: acknowledge, ask, attend, be aware, choose, describe, follow, give, hold, identify, listen, locate, name, receive, reply, select, etc.
Responding	▪ Active participation on the part of the student ▪ **Illustrative verbs:** agree, answer, ask, assist, communicate, comply, consent, conform, contribute, cooperate, discuss, follow-up, greet, help, indicate, inquire, label, obey, etc.
Valuing	▪ See worth or value in the subject, activity, assignment ▪ Individual's commitment to the underlying value guiding the behavior ▪ **Illustrative verbs:** accept, adopt, approve, complete, choose, commit, describe, desire, differentiate, display, endorse, exhibit, explain, express, form, initiate, justify, join, etc
Organization	▪ Bringing together a complex of values, possible, disparate values, resolving conflicts among them, and beginning to build an internally consistent value system ▪ **Illustrative verbs:** accept, adopt, approve, complete, choose, commit, describe, desire, differentiate, display, endorse, exhibit, explain, express, form, initiate, invite, work, etc
Characterization	▪ Internalization of values have a place in the individual's value hierarchy ▪ **Illustrative verbs:** act, advocate, behave, characterize, conform, continue, defend, devote, disclose, discriminate, display, encourage, endure, exemplify, etc.

C. Psychomotor Domain (Dave, R., 1967)

CATEGORIES	DESCRIPTIONS
Imitation	▪ Observe a skill and attempt to repeat it, or see a finished product and attempt to replicate it while attending to an exemplar. ▪ **Illustrative verbs:** begin, assemble, attempt, carry out, copy, calibrate, construct, dissect, duplicate, follow, mimic, move, practice, etc.
Manipulation	▪ Perform the skill or produce the product in a recognizable fashion by following general instructions rather than observation. ▪ **Illustrative verbs:** acquire, assemble, complete, conduct, do, execute, improve, maintain, make, manipulate, operate, pace, perform, produce, progress, use, etc.
Precision	▪ Independently perform the skill or produce the product, with accuracy, proportion, and exactness; at an expert level. ▪ **Illustrative verbs:** achieve, accomplish, advance, automatize, exceed, excel, master, reach, refine, succeed, surpass, transcend, etc.
Articulation	▪ Modify the skill or product the product to fit new situations; combine more than one skill in sequence with harmony and consistency. ▪ **Illustrative verbs:** adapt, alter, change, excel, rearrange, reorganize, revise, surpass, etc.
Naturalization	▪ Completion of one or more skills with ease and making the skill automatic with limited physical or mental exertion ▪ **Illustrative verbs:** arrange, combine, compose, construct, create, design, refine, transcend

-oOo-

TYPES OF TEST ACCORDING TO FORMAT

A. **Selective Type**- provides choices for the answer
   a. **Multiple Choice**- consists of a stem which describes the problem and 3 or more alternatives which give the suggested solutions. The incorrect alternatives are called **distractors**.
   b. **True-False or Alternate-choice**- consists of declarative statement that one has to mark true or false, right or wrong, correct or incorrect, yes or no, fact or opinion, and the like.

c. **Matching Type-** consists of two parallel columns: Column A, the column of **premises** from which a match is sought; Column B, the column of **responses** from which the selection is made. These questions help students see the relationships among a set of items and integrate knowledge.

B. **Supply test**- no choices are provided
   a. **Short-answer**- uses direct question that can be answered by a word, phrase, a number or a symbol
   b. **Completion test**- consists of an incomplete statement
   c. Essay test
      i. Restricted response- limits the content of the response by restricting the scope of the topic
      ii. Extended response- allows the student to select any factual information that they think is pertinent, to organize their answers in accordance with their best judgment.

Note: Generally, short-answer test provides the highest reliability, followed by multiple-choice, alternate-choice, and essay test.

Simple Guidelines for Writing Test Questions

Multiple-Choice Test Items

1. Avoid the tendency to make the correct answer longer than the distractors.
2. Using the same or similar words in both the stem and the correct answer can give away the answer.
3. Beware of grammatical giveaways. For example, if the stem ends with the word "an" and only one or two options begin with a vowel, then the student can easily eliminate the distractors.
4. Alert students can detect any tendency to prefer certain response options. For example, students may learn that option "c" is most often correct or that option "a" is seldom correct.
5. Avoid "*None of the above," "Some of the above," "All of the above*," phrases which usually scream out that they are the correct answer.
6. Order the response choices alphabetically, dates chronologically, formulas in terms of complexity. This logical sequence will help students locate choices.

Alternate-Choice Test / True-False Items

1. Base the item on a single idea.
2. Write items that test an important idea
3. Avoid lifting statements right from the textbook.
4. Make the questions a brief as possible
5. Write clearly true or clearly false statements. Write them in pairs: one "true" and one "false" version and choose one to keep balance on the test.
6. Eliminate giveaways:
   - Keep true and false statements approximately equal in length
   - Make half the statements true and half false.
   - Try to avoid such words as *"all," "always," "never," "only," "nothing,"* and *"alone."* Students know these words usually signify false statements.
7. Beware of words denoting indefinite degree. The use of words like "*more*," *"less," "important," "unimportant," "large," "small," "recent," "old," "tall,""great*," and so on, can easily lead to ambiguity.
8. State items positively. Negative statements may be difficult to interpret. This is especially true of statements using the double negative. If a negative word, such as "*not"* or "*never*," is used, be sure to underline or capitalize it.
9. Beware of detectable answer patterns. Students can pick out patterns such as (TTTTFFFF) which might be designed to make scoring easier.

Matching-Test Items

1. Provide directions. Students should not have to ask, for example, whether options may be used more than once.
2. Use only homogeneous material. Each item in a set should be the same as the other items, for examples all names or all numbers. When different kinds of items are used in each set, the associations tend to be obvious.
3. Place longer material in the left column. This will help students locate matches.
4. Arrange column material in some order. For example, names can be arranged alphabetically.
5. As a rule of thumb, the response set should contain a few more items than the premise set.
6. Keep the question to one page and on the same page. Arrange items so that students will not have to turn pages back and forth as they respond.

Completion-Test Items

1. Prepare a scoring key that contains all acceptable answers for each item.
2. Call for answers that can be scored objectively. Prefer single words or short phrases. Check your items by posing this question: Can someone with no competency in the subject score the items objectively by relying solely on the answer key?
3. Beware of open questions that invite unexpected but reasonable answers.

4. Eliminate giveaways.
    a. Make all the blanks an equal length
    b. Avoid grammatical clues such as "an."
5. Place the blanks near the end of the statement. Try to present a complete or nearly complete statement before calling for a response.
6. Limit the number of blanks to one or two per item. Statements with too many blanks waste time as students figure out what is being asked.
7. If a numerical answer is called for, indicate the units in which it is to be expressed.

Essay Questions

1. Use essay questions to assess complex learning outcomes.
2. Favor restricted-response essays that can be answered in about 15 minutes or less
3. Structure the problem. This will make it easier to grade.
4. Prepare model answers before asking students to respond
5. Allow sufficient time to answer to give the students to outline first.
6. Encourage thoughtful answers by writing positive and constructive comments
7. Require all students to answer the same questions

Guidelines for Scoring Essay Questions

1. Use model answers.
2. Score the same question on all papers before going to the next question.
3. Cover student names. This will reduce the likelihood of biased scoring
4. Read each essay twice before scoring.
5. Have other colleagues independently rate the students' response. Getting the average of all the ratings will increase the reliability of the test score regardless of scoring technique.

ALTERNATIVE ASSESSMENT

A. PERFORMANCE-BASED ASSESSMENT- a mode of assessment that requires actual demonstration of skills or creation of products of learning.

Authentic assessment refers to the performance-based assessment that is valued on its own right for being the real instances of extended criterion performance and this involves application of a skill beyond the instructional context.

Note: All authentic assessments are performance-based but not all performance-based assessments are authentic assessments.

WHEN TO USE PERFORMANCE-BASED ASSESSMENT	ADVANTAGES	LIMITATIONS
▪ Specific behaviors or behavioral outcomes are to be observed ▪ Possibility of judging the appropriateness of students' actions ▪ A process or outcome cannot be directly measured by paper-and-pen test	▪ Allow evaluation of complex skills which are difficult to assess using written tests ▪ Positive effect on instruction and learning ▪ Can be used to evaluate both the process and the product	▪ Time-consuming to administer, develop and score ▪ Urgent scoring ▪ Subjectivity in scoring Inconsistencies in performance on alternative skills

CATEGORIES OF PERFORMANCE-BASED ASSESSMENT

PROCESS vs. PRODUCT	
Process refers to the *procedure* that a student uses to complete a task.	Product is the tangible *outcome* that may be the result of completing a process
SIMULATED vs. REAL SETTING	
Simulated setting refers to the alternative setting that can adequately *substitute the real thing*	Real setting refers to the *authentic venue* where learners will perform the task
STRUCTURED vs. NATURAL STIMULI	
Structured stimuli is used to ensure that a performance/behavior to be observed will occur. This tends to elicit *maximum* performance.	Natural stimuli requires no intervention of the observer. This tends to elicit *typical* performance.

B. PORTFOLIO ASSESSMENT

Portfolio- meaningful collection of student work that exemplifies interest, attitude, ranges of skills, and development over a period of time. **Professional portfolio** play a summative role whereas **student portfolios** are used for formative purposes.

Characteristics:

1. Adaptable to individualized instructional goals
2. Focus on assessment of products
3. Identify students' strengths rather than weaknesses
4. Actively involve students in the evaluation process
5. Communicate student achievement to others
6. Time-intensive
7. Reliability

TYPES OF PORTFOLIO

TYPES	DESCRIPTION
Showcase	▪ A collection of students' best work
Reflective	▪ Used for helping teachers, students, and family members think about various dimensions of student learning such as effort, achievement, etc
Cumulative	▪ A collection of items done for an extended period of time ▪ Analyzed to verify changes in the products and process associated with student learning
Goal-based	▪ A collection of works chosen by students and teachers to match pre-established objectives
Process	▪ A way of documenting the steps and processes a student has done to complete a piece of work

RUBRICS

Rubrics is a scoring guides, consisting of specific pre-established performance criteria, used in evaluating student work on performance assessments.

TYPES OF RUBRICS:

1. **Holistic Rubric**- requires the teacher to score the overall process or product as a whole, without judging the component parts separately. The focus of the score reported using this type of rubric is on the overall quality; however, only limited feedback is provided to the student. This is more appropriate to use when there is no definite correct answer, there is a need to quickly score a performance or a summative score is desired.
2. **Analytic Rubric**- requires the teacher to score individual components of the product or performance first, then sums the individual scores to obtain a total score. This is preferred when there may be one or two acceptable responses, creativity is not an essential feature of the students' response, or students need formative feedback.

STEPS IN DESIGNING SCORING RUBRICS(Raagas, 2010)**:**

1. Re-examine the learning objectives to be addressed by the task.
2. Identify specific observable attributes that you want to see (as well as those you don't want to see) your students demonstrate in their product, process, or performance.
3. Brainstorm characteristics that describe each attribute.

4a. For holistic rubrics, write thorough narrative descriptions for excellent work and poor work incorporating each attribute into the description

4b. For analytic rubrics, write thorough narrative descriptions for excellent work and poor work for each individual attribute.

5a. For holistic rubrics, complete the rubric by describing other levels on the continuum that ranges from excellent to poor work for the collective attributes.

5b. For analytic rubric, complete the rubric by describing other levels on the continuum that ranges from excellent to poor work for each attribute.

6. Collect samples of student work that exemplify each level.
7. Revise the rubric, as necessary.

BIASES AND SCORING ERRORS

1. Halo effect- letting general impression of student influence rating of specific criteria
2. Contamination effect- judgment is influenced by irrelevant knowledge about the student or other factors that have no bearing on performance level
3. Similar-to-me-effect- judging favorably those students whom faculty see as similar to themselves
4. Contrast effect- judging by comparing student against other students instead of established criteria and standards
5. Rater drift- Unintentionally redefining criteria and standards over time or across a series of scorings
6. Leniency error- tendency to judge a performance or product better than it really is
7. Generosity error- tendency to use high end of scale only
8. Severity error- tendency to use low end of scale only

9. Central tendency error- tendency to avoid both extremes of the scale

C. AFFECTIVE ASSESSMENT

COMMON AFFECTIVE TARGETS	DESCRIPTIONS
Attitude	▪ Mental states that structure the way a student perceives his environment and this guides the way he responds to it
Interest	▪ A disposition organized through experience which impels an individual to seek out particular activities or objects for attention or acquisition
Value	▪ Set of beliefs on what is desirable and not
Academic self-concept	▪ Sum of all evaluative judgments one makes about one's possibility of success or productivity in an academic context
Self-esteem	▪ Beliefs on what a person can and cannot do
Locus of control	▪ This refers to views/belief on the reason for one's success or failure. It can be internal or external.

METHODS	DESCRIPTIONS and TOOLS
Teachers observation	▪ Can be structured or unstructured. Inferences are made from what was observed ▪ Tools/Techniques: ○ **Anecdotal Records**- a narrative record of observations of a particular learner behavior in an unstructured setting ○ **Clinical Observation**- this involves use of sophisticated instruments in a controlled setting ○ **Scales**- consists of list of characteristics or behaviors to be observed and an evaluative scale to indicate the degree to which they occur ○ **Checklist**- a set of traits that an observer has to mark if demonstrated by a particular learner
Peer ratings	▪ Least common among the three methods due to the relatively inefficient nature of conducting, scoring, and interpreting peer ratings ▪ Tools/Techniques: ○ **Sociometric technique**- provides information on social relationships such as degrees of acceptance, roles and interactions within groups ○ **Guess Who Technique**- best used with relatively well-established groups in which members are well acquainted with each other ○ **Communigram**- assesses the frequency of verbal participation of a learner in a particular group within a given period ○ **Social Distance Scales**- measures the distance of a learner between other persons and himself that is usually identified through the reaction to given statements that compare attitudes of acceptance or rejection of other people
Students self-reports	▪ Includes interviews, questionnaires and surveys ▪ Tools/Techniques: ○ **Autobiography**- enables the learners to describe his/her own life and experiences ○ **Self-Expression Essay**- seeks to assess the learner's response to a particular question or concern usually in a short written essay form ○ **Self-Description**- requires the learner to paint a picture of himself/herself in words ○ **Self-Awareness Exercises**- designed to help learners become more aware of their feelings, emotions, and values ○ **Questionnaire**- provides an opportunity to easily collect a great deal of information that may be useful in understanding the learner ○ **Structured Interview**- enables the counselor to obtain specific information and to explore in-depth behavior or responses

ATTITUDINAL SCALE

A. **Closed-Item or Forced-choice instruments**- answers are selected from the given choices
   1. **Checklist**- measures students preferences, hobbies, etc. by marking a set of possible responses
   2. **Ranking**- measures student's preferences or priorities by ranking a set of attitudes or objects
   3. **Scales**- indicates the extent or degree of one's response
      a. **Rating Scale**- measures the degree or extent of one's attitude, feelings, and perception towards objects, people, and ideas by making a point along 3 to 5-point scale.

b. **Likert Scale**- measures the degree of one's agreement or disagreement to a variety of statements related to student's attitude
c. **Semantic Differential Scale**- measures the degree of one's attitude, , feelings, and perception towards objects, people, and ideas by making a point along 5 or 7 or 11-point scale of semantic objectives
d. **Stapel Scale**-simplified versions of semantic-differential scales which use only one pole. Students rate the object/issue/idea by selecting a numerical category.

B. **Open-Ended Instruments**- no choices are provided
1. **Sentence Completion**- measures students preferences over a variety of attitudes and allows students to answer by completing an unfinished statement which may vary in length
2. **Surveys**- measures the values held by an individual by writing one or many responses to a given question
3. **Essays**- allows the students to reveal and clarify their preferences, hobbies, attitudes, feeelings, beliefs, and interests by writing their reactions or opinions to a given question

-oOo-

VALIDITY AND RELIABILITY

VALIDITY

This refers to the degree to which a test measures what is intended to be measured. It is the usefulness of the test for a given purpose. It is the most important criterion of a good examination.

WAYS TO ESTABLISH VALIDITY

1. **Face validity**- this is done by examining the physical appearance of the test. However, "face validity" is considered a misnomer because it addresses appearance rather than determining if a test measures what it intends to measure.
2. **Construct validity**- establishes a link between the underlying psychological construct we wish to measure and the visible performance we choose to observe
3. **Content-related evidence of validity**- done through a careful and critical examination of the objectives of the test so that it reflects the curricular objectives
4. **Criterion-related evidence of validity**- indicates how well performance on a test correlates with performance on relevant criterion measures external to the test.

Table of Specifications- consists of a two-dimensional chart. The vertical dimension of the chart lists the content area to be addressed by the test and the horizontal dimension lists the categories of performance the test is to measure. The numbers within the table of specifications indicate the number of test questions to be associated with each content area and capability.
Note: Constructing a **Table of Specifications** establishes **content validity** NOT construct validity.

RELIABILITY

This refers to the consistency of scores obtained by the same person when retested using the same instrument or one that is parallel to it

FACTORS AFFECTING RELIABILITY (SOURCES OF ERRORS)

1. *Item Sampling* —any test is only a sample of all possible items, the item sample itself can be a source of error.Longer tests are typically more reliable because we get a better sample of the course content and students' performance.
2. *Con*struction of the Items — Major threat to reliable measurement is poorly worded or ambiguous questions or trick questions.
3. *Test administration*–Environmental factors such as heat, light, noise, confusing directions, and different testing time allowed to different students can affect students' scores,
4. *Scoring — Objectivity* or the extent to which equally competent scores obtain the same score is a factor affecting reliability
5. *Difficulty of the Test*–A test that is either too easy or too difficult for the class taking it will typically have low reliability.
6. *Student Factors*–Student fatigue, illness, or anxiety can induce error and lower reliability because they affect performance and keep a test from being a measure of their true ability or achievement.

METHOD	TYPE OF RELIABILITY MEASURE	PROCEDURE	STATISTICAL MEASURE
Test-retest	Measure of stability	▪ Give a test twice to the same group with any time interval	Pearson r

		between sets from several minutes to several years	
Equivalent forms	Measure of equivalence	▪ Give parallel forms of test at the same	Pearson r
Test-retest with Equivalent forms	Measure of stability and equivalence	▪ Give parallel forms of test with increased time interval between forms	Pearson r
Split Half	Measure of Internal Consistency	▪ Give a test once. Score equivalent halves of the test.	Pearson r **and** Spearman-Brown Formula
Kuder-Richardson (Used for 0-1 test)		▪ Give the test once, then correlate the proportion/percentage of the students who got the item correctly	K-R20 **or**K-R21
Cronbach Coefficient Alpha (Used for Essay test)		▪ Give a test once. Then estimate reliability by using the standard deviation per item and the standard deviation of the test scores.	Cronbach Coefficient Alpha Formula

Note: For the Kuder-Richardson Method, **use K-R20**if the proportion of correct responses to a particular item does NOT vary widely.

Interpretation of the Correlation Coefficient

1	-------	Perfect Positive Correlation
		High positive correlation
0.5	-------	Positive Correlation
		Low Positive correlation
0	-------	No relationship
		Low negative correlation
-0.5	-------	Negative correlation
		High Negative correlation
-1	-------	Perfect Negative Correlation

-oOo-

ITEM ANALYSIS

STEPS:

1. Score the test. Arrange the scores from highest to lowest.
2. Get the top 27% (upper group) and below 27% (lower group) of the examinees.
   27% is called the criterion group.
3. Count the number of examinees in the upper and lower group who got each item correctly.
4. Compute the Difficulty Index of each item:

$$Df = \frac{Pt + Pb}{2n} \quad where\ n = no.of\ students\ in\ each\ group$$

5. Compute for the Discrimination Index

$$Df = \frac{Pt - Pb}{n}$$

6. Compute for the Index of Effectiveness of the Distractors (for Multiple-Choice Items)

$$IE = \frac{Hg - Lg}{n}$$

Note: If there are 30 students or less, the criterion group is 50% of the class instead of 27%.

INTERPRETATIONS

Difficulty Index Scale	DISCRIMINATION INDEX	INDEX OF EFFECTIVENESS
Below 0.25 -Difficult (reject, revise or retain)	0.40 and above- very good item (retain)	Below 0- plausible (retain)
0.25-0.75 -Average (revise or retain)	0.30-0.39 - reasonably good (retain or revise)	Above 0- not working well (revise)
Above 0.75 –Easy	0.20 to 0.29 - marginal item	0- not plausible

(reject, revise or retain)	(retain or revise)	(reject or revise)*
	Below 0.20 - poor item (reject or revise)	

***** Reject if no one chose the option, revise if the no. of students in the upper and lower group who chose the option is equal.

-oOo-

FREQUENCY POLYGON

CHARACTERISTIC	CLASSIFICATION and DESCRIPTION
Symmetry (See Figure 1b)	**Symmetric-** has two halves which mirrors each other. The value of the skewness is zero. **Nonsymmetric or Asymmetric-** skewed curve
Skewness (See Figure 1a and 1c)	**Positively skewed (skewed to the RIGHT)-** most scores are below the mean and there are extremely low scores. In this distribution, typically, the mode is lower than the median and the median is lower than mean. **Negatively skewed (skewed to the LEFT)-** most scores are above the mean and there are extremely low scores. The mean is lower than the median and the median is lower than the mode.
Modality	**Unimodal**- curve with 1 mode **Bimodal**- curve with 2 modes **Multimodal/Polymodal**- curve with 3 or more modes **Rectangular Distribution**- mode does not exist
Kurtosis (See Figure 2)	**Leptokurtic-** highly peaked and the tails are more elevated above the baseline **Mesokurtic**- moderately peaked **Platykurtic**- flattened peak

Note: In classifying a frequency polygon in terms of skewness, base it on the direction of the tail and not on location of the hump or peak.

Figure 1

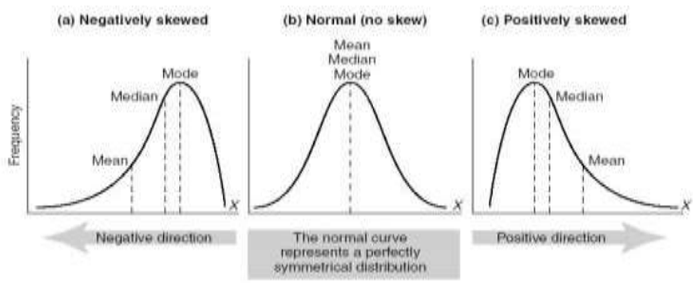
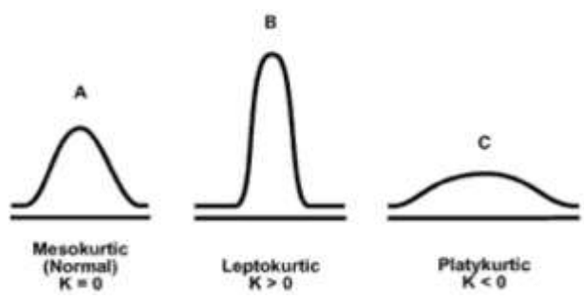


Figure 2



MEASUREMENT SCALES

MEASUREMENT	CHARACTERISTICS	EXAMPLES
Nominal	▪ A scale that measures data by name only	Gender, Religion
Ordinal	▪ Rank data ▪ Distance between points are indefinite	Income, letter grades
Interval	▪ Distance between points are equal ▪ No absolute zero	Test scores Temperature
Ratio	▪ Has absolute zero	Height Weight

MEASURES OF CENTRAL TENDENCY AND VARIABILITY

ASSUMPTIONS WHEN USED	APPROPRIATE STATISTICAL TOOLS	
	Measures of Central Tendency (describes the representative value of a set of data)	**Measures of Variability/Dispersion** (describes the degree of spread or dispersion of a set of data)
▪ When the frequency distribution	**Mean**	**Standard Deviation**

is regular or symmetric (bell-curved) ▪ Usually used when data are numeric (**interval or ratio**)	▪ arithmetic average ▪ sensitive to extreme scores	▪ The root-mean-square of the deviations from the mean
▪ When the frequency distribution is irregular or skewed ▪ Usually used when the data is **ordinal, interval or ratio**	**Median** ▪ middle most score in a group of scores in an array	**Quartile deviation** ▪ The average deviation of the 1st and 3rd quartiles from the median
▪ When quick answer is needed ▪ Usually used when the data are **nominal or ordinal**	**Mode** ▪ most frequently occurring score	**Range** ▪ The difference between the highest and the lowest score in thedistribution

Application:

Given the scores: 20, 20, 21, 24, 25, 26, 27, 27, 27, 31. Solve for the measures of central tendency.

Solution:

1. Mean = sum of all the scores ÷ by the no. of scores
   = (20+20+21+24+25+26+27+27+27+31) ÷10 = 248 ÷ 10 = 24.8
   Hence, the mean of the given scores is 24.8
2. Median= middlemost score (get the average of 25 and 26) = (25+26) ÷2 = 25.5
   Hence, the median is 25.5
3. Mode= most frequently occurring score = 27 (This score repeats 3 times)

Since the mean, median and mode is 24.8, 25.5 and 27, the distribution of scores must be negatively skewed or skewed to the left.

STANDARD SCORES

The major purpose of standard scores is to place scores for any individual on any variable having any mean and standard deviation on the same standard scale so that comparisons can be made. Without some standard scale, comparisons across individuals and/or across variables would be difficult to make (Lomax,2001).

A. **Percentile Score**- tells the percentage of examinees that lies below one's score
   Formula: $P_i=i(n+1)/100^{th}$ place
B. **Z-Score**- tells the number of standard deviations equivalent to a given raw score
   Formula: Z= (score- mean)/standard deviation
C. **T-score**- refers to any set of normally distributed standard deviation score that has a mean of 50 and a standard deviation of 10
   Formula: T-Score= 50+10(Z)
D. **Stanine Score**- a standard score with a mean equal to 5 and standard deviation of 2.
   Formula: 5+2(Z)

Application

Example1 : Given the scores: 14, 35, 45, 55, 55, 56, 56, 65, 87, 89, 92, find the 85th percentile.

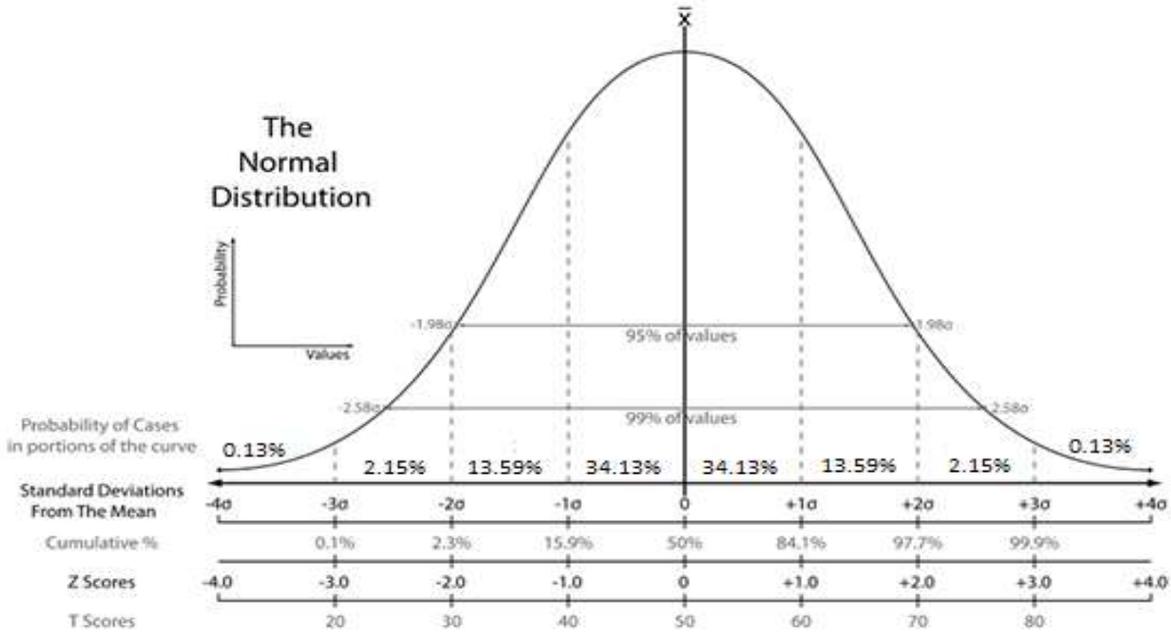
Solution: P85= 85(11+1)/100th place=85(12)/100th place=1020/100th place=10.2th place
Look for the score in the 10.2th place (between 10 and 11th place).
P85= 10th place + 0.2(11th -10th place)= 89+0.2(92-89) = 89 + 0.2(3) = 89+0.6= 89.6

Therefore, the 85th percentile is 89.6. This means that 85% of the examinees who took the same exam got a score lower than 89.6. Conversely, 15% of the examinees got a score of at least 89.6.

Example2: Mico's raw score in the quarter exam is 50. If the mean of the class is 45 and the standard deviation is 2.5, solve for the Z score, T-score and Stanine score of Mico.

Solution: **Z-score**= (score-mean)/SD = (50-45)/2.5 = (5)/2.5 = 2
Hence, Z= 2 (Use this value to solve for the T-score and Stanine Score)
T-score = 50 + 10 (Z) = 50 + 10 (2) = 50 + 20 = 70
Stanine score = 5 + 2(Z) = 5 + 2(2) = 5 + 4 = 9

Therefore, Mico's score in the quarter exam is 2 standard deviation above the mean. This implies that his score is higher than 97.72% of that of his classmates. Further, his performance can be classified as above average.



ASSIGNING GRADES/MARKS/RATINGS

Marking or grading is the process of assigning value to a performance.

Marks/Grades/Rating Symbols:

A. Could be in:
   - Percent such as 70%, 88%, 95%
   - Letters such as A, A-, B, B-, C, D
   - Numbers such as 1.0, 1.5, 2.75, 5
   - Descriptive expressions such as Outstanding (O), Very Satisfactory (VS), Satisfactory (S), Moderately Satisfactory (MS), Needs Improvement (NI)

B. Could represent:
   - How a student is performing in relation to other students (norm-referenced grading)
   - The extent to which a student has mastered a particular body of knowledge (criterion-referenced grading)
   - How a student is performing in relation to a teacher's judgment of his or her potential

C. Could be for:
   - **Certification** that gives assurance that a student has mastered a specific content or achieved a certain level of accomplishment
   - **Selection** that provides basis in identifying or grouping students for certain educational paths or programs
   - **Direction** that provides information for diagnosis and planning
   - **Motivation** that emphasizes specific material or skills to be learned and helping students to understand and improve their performance

D. Could be based on:
   - Examination results or test data
   - Observation of students works
   - Group evaluation activities
   - Class discussion and recitations
   - Homeworks
   - Notebooks and note taking
   - Reports, themes and research papers
   - Discussions and debates
   - Portfolios
   - Projects
   - Attitudes, etc.

E. Could be assigned by using:
   - **Criterion-referenced grading**- grade is based on fixed or absolute standards where grade is assigned on how student has met the criteria or well-defined objectives of a course that were spelled out.
   - **Norm-referenced grading**- grade is based on relative standards where a student's grade reflects his or her level of achievement relative to the performance of the other students in class.
   - **Point or percentage grading system**- the teacher identifies points or percentages for various tests and class activities depending on their importance. The total of these points will be the bases for the grade assigned to the student.

- **Contract grading system**- each student agrees to work for a particular grade according to agreed- upon standards.

GUIDELINES IN GRADING STUDENTS

1. Explain your grading system to the students early in the course and remind them of the grading policies regularly.
2. Base grades on a predetermined and reasonable set of standards.
3. Base you grades on as much objective evidence as possible.
4. Base grades on the student's attitude as well as achievement, especially at the elementary and high school level.
5. Base grades on the student's relative standing compared to classmates.
6. Base grades on a variety of sources.
7. As a rule, do not change grades, once computed.
8. Become familiar with the grading policy of your school and with your colleague's standards
9. When failing a student, closely follow school procedures.
10. Record grades on report cards and cumulative records.
11. Guard against bias in grading.
12. Keep pupils informed of their standing in the class.