

#### 4.1 Apply principles in constructing traditional and alternative/authentic forms of high quality assessment

1. Here is an essay test for the finals given by a teacher to his class: Write everything that you learned from this course.

1.1 What's wrong with the test item?

- I. It is too broad.
- II. It does not show how scoring will be done.
- III. It is very open-ended.

- A. I, II and III
- B. II and III
- C. I and II
- D. I only

2. Here is an essay test for the finals given by a teacher to his class: Write everything that you learned from this course on school-based management. Which is an improved version of the essay test?

- A. Write 5 main principles on school-based management that you learned from the course. You will be given 2 points each for stating a correct principle and 8 points each for correctly and adequately explaining each principle.
- B. State and explain 5 main principles of school-based management.
- C. Clearly state and fully explain 5 main principles of school-based management.
- D. State and explain 5 main principles of school-based management. Five points for stating and 25 points for correctly explaining.

3. Here are 4 elements with their corresponding percentages that the K to 12 Curriculum include in its assessment process:

Knowledge – 15%    Processes and Skills- 25%    Understandings -30%    Product and Performance -30%

Which conclusion can be derived from the K to 12 assessment practice?

- A. Assessment puts emphasis on knowledge and processes and skills.
- B. Assessment practice makes use of multiple sources.
- C. Assessment emphasizes most on comprehension.
- D. Assessment requires critical thinking and disregards recall.

4. Here are 4 elements with their corresponding percentages that the K to 12 Curriculum include in its assessment process:

Knowledge – 15%    Processes and Skills- 25%    Understandings -30%    Product and Performance -30%

What generalization on assessment can be derived from the Table?

- A. Assessment considers recall of information insignificant.
  - B. Assessment in the level of understanding excludes learning in the knowledge level.
  - C. Assessment considers demonstration of a skill learned as the strongest proof of learning.
  - D. Assessment puts more emphasis on knowledge and processes and skills.
5. Why is it considered unfair to learners for teachers to use only the traditional tools for assessment?
- A. Learners learn and demonstrate learning in multiple ways other than in written form.
  - B. Traditional tools cannot measure higher order thinking skills.
  - C. Most teachers are not ready to use alternative assessment tools.
  - D. Traditional assessment is threatening.
6. Which best explains assessment as learning?
- I. The use of a scoring rubric makes specific to the learner the ideal expected output.
  - II. The use of a scoring rubric gives the learner an opportunity to for self-assessment.
  - III. A teacher who periodically assesses learners' understanding ensures learning.
- A. I and II
  - B. II only
  - C. I only
  - D. I, II and III
7. Which is referred to as assessment of learning?
- A. Formative assessment
  - B. Summative assessment
  - C. Traditional assessment
  - D. Authentic assessment
8. Which applies to authentic assessment?
- A. Indirect assessment
  - B. Teaching to the test
  - C. Performing real-world task
  - D. Paper-and-pencil test
9. Which applies to authentic assessment?
- A. Product assessment
  - B. Indirect assessment

- C. Traditional test
  - D. Teaching to the test
10. Which applies to authentic assessment?
- A. Application of knowledge and skills learned
  - B. Teaching to the test
  - C. Indirect assessment
  - D. Traditional test
11. Which go/es with authentic assessment?
- A. Task to perform and grading system
  - B. Test items to answer and scoring system
  - C. Task to perform and scoring rubric
  - D. Test items to answer and grading system
12. Which attribute belongs more to traditional assessment?
- A. Performing a task
  - B. Selecting a response
  - C. Direct evidence
  - D. Real –life
13. Which attribute belongs more to traditional assessment?
- A. Indirect evidence
  - B. Performing a task
  - C. Direct evidence
  - D. Real –life
14. Which attribute belongs more to traditional assessment?
- A. Performing a task
  - B. Direct evidence
  - C. Real–life
  - D. Contrived incident
15. Which goes with authentic assessment?
- A. Application
  - B. Contrived incident
  - C. Simple recall
  - D. Indirect evidence

16. Which goes with authentic assessment?
- A. Direct evidence
  - B. Contrived incident
  - C. Simple recall
  - D. Indirect evidence
17. Which goes with authentic assessment?
- A. Contrived incident
  - B. Simple recall
  - C. Indirect evidence
  - D. Performing a task
18. Between a driver who passed the written test but failed the driving test and one who passed the driving test but failed the written test, the latter is generally preferred. What does this imply about assessment practice?
- A. Authentic assessment is preferred to traditional assessment.
  - B. Traditional assessment is more reliable than authentic assessment.
  - C. Traditional assessment supplements authentic assessment.
  - D. Traditional and authentic assessments cannot adequately measure what is intended to be measured.
19. Between a driver who passed the written test but failed the driving test and one who passed the driving test but failed the written test, the latter is generally preferred. But most preferred is one who passed both written and driving test. What does this imply about assessment practice?
- A. Traditional and authentic assessments cannot adequately measure what is intended to be measured.
  - B. Authentic assessment complements traditional assessment.
  - C. Authentic assessment can be relied upon.
  - D. Traditional assessment when done well can function as authentic assessment.
20. Is traditional assessment grounded on the following philosophies?
- I. A school's mission is to develop productive citizens.
  - II. To be a productive citizen, an individual must possess a certain body of knowledge and skills.
  - III. Schools must teach this body of knowledge and skills and test if students acquired these knowledge and skills.
- A. No.

- B. Yes, if productivity of a citizen is tested by real-world tasks.
- C. Yes.
- D. No, except I.

21. On what philosophy is authentic assessment anchored?

Schools must help students :

- A. become proficient at performing tasks they will encounter when they graduate.
- B. master knowledge and skills.
- C. pass tests on knowledge and skills acquired.
- D. acquire written and communication skills.

For Questions 22-27. Here are two scoring rubrics. Study them then answer the following questions:

Rubric A

Criteria		1	2	3
Number of Sources	X 1	1-4	5-9	10-12
Historical accuracy	X3	Lots of historical inaccuracies	Few inaccuracies	No apparent
Organization	X1	Cannot tell from which source information came	Can tell with difficulty where information came from	Can easily tell from which sources of info was drawn
Bibliography	X1	Bibliography contains very little information	Bibliography contains most relevant information	All relevant information is included

Rubric B

Criteria		Poor (1)	Good (2)	Excellent (3)
Number of Sources	X 1			
Historical accuracy	X 3			

Organization X 1

Bibliography X 1

22. For which are these scoring rubrics?
- A. Research portion of a project
  - B. A project in a History class
  - C. An essay in an English class
  - D. Theme written in a composition
23. How are Rubrics A and B classified?
- A. Holistic rubric
  - B. Analytic rubric
  - C. Values rubric
  - D. Skills rubric
24. Rubrics A and B are meant to assess research portion of a project. Against how many criteria is the research assessed?
- A. 4
  - B. 5
  - C. 6
  - D. 3
25. Are the criteria for research given equal weight?
- A. Yes.
  - B. No, organization is given most weight
  - C. No, historical accuracy is given most weight.
  - D. Yes, for two criteria.
26. Among the criteria for good research, which one is given greatest importance?
- A. Number of sources
  - B. Organization
  - C. Historical accuracy
  - D. Bibliography
27. What are in Rubric A that are not in Rubric B?
- A. Specific descriptors
  - B. Criteria
  - C. Levels of performance
  - D. Weights

For Questions 28 - 30

Scoring Rubric C

3- Excellent researcher

- Included 10-12 sources
- No apparent historical inaccuracies
- Can easily tell from which resources information was drawn
- All relevant information is included

2- Good Researcher

- Included 5-9 sources
- Few historical inaccuracies
- Can tell with difficulty where information came from
- Bibliography contains most relevant information

1-Poor Researcher

- Included 1-4 sources
- Lots of historical inaccuracies
- Cannot tell from which source information came
- Bibliography contains very little information

28. Under what type of Scoring Rubric does C fall?

- A. Analytic  
B. Descriptive  
C. Holistic  
D. Quantitative

29. How many levels of performance are there in Scoring Rubric C?

- A. Three  
B. Depends on the research scored  
C. Four  
D. Two

30. When is holistic rubric preferred?

- I. When gross judgment on a piece of work needs to be made  
II. When one wants to assess each criterion separately  
III. When one wants a global picture of performance  
A. I and III  
B. I only  
C. I and II  
D. III only

31. For diagnostic purposes, which Scoring Rubric will be most appropriate?

- A. Analytic  
B. Holistic  
C. Global  
D. Descriptive

32. Complete the analogy: Traditional assessment: selecting a response  
Authentic assessment: \_\_\_\_\_

- A. performing a task  
B. selecting a response  
C. filling in the blanks  
D. selecting a response

- B. matching the answer                      D. identifying the answer

33. Complete the analogy: Authentic response: real-life  
Traditional assessment: \_\_\_\_\_  
A. real task                                      C. contrived  
B. simulated activity                          D. role-playing

34. Complete the analogy : Traditional assessment : recall  
Authentic assessment : \_\_\_\_\_  
A. comprehension                              C. recognition  
B. interpretation                                D. application

35. Complete the analogy : Traditional assessment : indirect evidence  
Authentic assessment : \_\_\_\_\_  
A. circumstantial evidence                  C. direct evidence  
B. conclusion                                    D. generalization

36. Authentic assessments are said to be more student –structured than teacher-structured. What does this mean?

Authentic assessment:

- A. allows more student choice in determining what is presented as evidence of proficiency.
- B. is limited to the presentation of evidence of learning that is designed by the teacher
- C. limits presentation of evidence of learning to that which is acceptable to the teacher
- D. allows students to choose the topic/s to learn over which to show proof of learning

37. Why are traditional assessments said to be more teacher-structured than student structured?

- A. What students will show as proof of learning is dependent on what teacher asks.
- B. Students can show evidence of learning much more than what teacher asks.
- C. In traditional assessment, teachers limit tests to low-level test items.
- D. Traditional assessment is not as challenging as authentic assessment.

38. Are authentic assessments more student – structured than teacher-structured?

- A. No, it is more teacher-structured.
  - B. It depends on the teacher giving the authentic assessment .
  - C. Yes. is more student-structured.
  - D. It depends on the type of authentic assessment given.
39. Is traditional assessment more student- structured than teacher-structured?
- A. No, it is more teacher-structured.
  - B. It depends on the orientation of the teacher giving the traditional assessment.
  - C. Yes, it is more student-structured.
  - D. It can be determined by the ability of the students taking the test.
40. What does “teaching to the test” mean?
- A. Teachers teach by focusing primarily on those that will be tested.
  - B. Teachers teach in such a way that students perform well on meaningful tasks.
  - C. Teachers use test to teach.
  - D. Teachers begin and end instruction with a test.
41. Which one shows that authentic assessment encourages integration of teaching, learning and assessing?
- A. Students are asked to give back information they were taught.
  - B. Students are asked to apply what they were taught.
  - C. Students learn when they solve real-world problem, assess their own solutions to the problem with a teacher facilitating.
  - D. Teacher presents a word problem in Math, gives formula to solve the problem and asks students to check on their answers.
42. Assessments should be multiple and varied. What does this mean?
- I. Sufficient number of samples of student performance must be obtained.
  - II. Use different tools to determine students’ application of learning
  - III. Cater to the multiple intelligences
- A. I, II
  - B. II and III
  - C. I, II
  - D. I, II, and III
43. Which illustrate/s multiple and varied assessment?
- Use observation-based assessment tool and:
- I. observe not only one but all the students.
  - II. observe as frequently and as regularly as possible.
  - III. stick to only one observation tool





- B. Repeated options
- C. No answer
- D. Poor grammar

53. Read the test item then answer the question:

*Internal consistency is high...*

- A. *when students who scored high on the first half of the test score high on the second half of the test.*
- B. *when students who scored high on the first half of the test score low on the second half of the test.*
- C. *when students who scored high on the first half of the test score in an unpredictable manner on the second half of the test.*
- D. *when all of the above are true.*

**What is/are faulty with the test item?**

- I. It uses “all of the above “.
- II. Not all options are grammatically correct when connected to the stem.
- III. The options are repetitive.

- A. I and II
- B. II and III
- C. I only
- D. I and III

54. Is this an improved version of the test item given in # 53?

*If the internal consistency of a test is good, how will a group of students score on the second half of the test if they got the highest scores on the first half of the test?*

- A. *Highest scores*
- B. *Lowest scores*
- C. *Unpredictable scores*
- D. *Cannot be determined*

- A. No, the number of options is reduced to 3.
- B. Yes, the options are not repetitive.
- C. All options implausible.
- D. Yes, there is window dressing.

55. Read this question then answer [test item](#):

*Various item formats have specific advantages and limitations. An advantage the essay format has over the multiple-choice format is:*

- A. *the essay item can assess more skills in a given amount of time.*
- B. *the essay item can assess students' ability to evaluate ideas.*
- C. *the essay item can be reliably scored.*
- D. *the essay item requires students to communicate ideas in writing.*

What is TRUE of the test item?

- A. **It is poorly written because it has window dressing.**
- B. It is well written.
- C. It has two possible answers
- D. The options end with a period.

56. Read the test item then answer the question:

*Which of the following has helped most to increase the length of human life?*

- A. *Fast driving*
- B. *Avoidance of overeating*
- C. *Wider use of vitamins*
- D. *Wider use of inoculation*

What is faulty with the test item ?

- A. **Implausible distracter**
- B. No stem
- C, Options not parallel
- D. Options grammatically inconsistent with the stem

57. Read the test item then answer the question:

*In purifying water for a city water supply, one process is to have the impure water seep through layers of sand and fine and coarse gravel. Here many impurities are left behind. Below are four terms, one of which will describe this process better than the others. Select the correct one.*

- A. *Sedimentation*
- B. *Filtration*
- C. *Chlorination*
- D. *Aeration*

What is a flaw of the test item?

- A. Stem includes an “instructional aside.”
- B. The test item is very long.
- C. There is no correct answer.
- D. There are two possible answers.

58. Read the test item then answer the question:

*Which event is **most** important in Philippine history?*

- A. *The capture of Aguinaldo*
- B. *The execution of Jose Rizal*
- C. *EDSA I Revolution*
- D. *Proclamation of Philippine Independence*

What is a flaw of the test item ?

- A. **Opinionated**
- B. No single answer
- C. Insignificant
- D. Irrelevant

59. Read the test item then answer the question:

*The population of Denmark is about*

- A. *2 million.*
- B. *15 million.*
- C. *4 million.*
- D. *7 million.*

What is a flaw of the test item?

- A. **Options not sequenced**
- B. Insignificant
- C. Irrelevant
- D. Limited options

60. In which level in Bloom’s cognitive taxonomy of objectives are questions that begin with

“restate in your own words...” and “convert fractions into... fall?”

- A. Analysis
- B. **Comprehension**
- C. Recall
- D. Application

61. Which questions fall under recall in Bloom’s taxonomy?

- I. Who invented the...?
- II What is meant by...?

- III. Where is the...?
- IV. Restate in your own words ...

- A. I and II
- B. II and IV
- C. I and III
- D. I, II and III

62. Questions that begin with

- How is...an example of... ?
- How is...related to... ?

in Bloom's taxonomy.

Are questions on \_\_\_\_\_.

- A. analysis
- B. synthesis
- C. comprehension
- D. application

63. Questions such as:

- What would you infer from... ?
- What ideas can you add to... ?
- How would you create a...
- are examples of \_\_\_\_\_ in Bloom's and Anderson's taxonomy of objectives

- A. Creating
- B. Analyzing
- C. Applying
- D. Understanding

64. In presenting words and their definitions, in a matching type of test, which practice is CORRECT?

- A. Column 2 should contain the definitions, which are longer than the words
- B. The items which are arranged in chronological order would be found in both Columns.
- C. The premise is found in Column 1.
- D. All the items in both columns are numbered.

65. Which type of paper-and-pencil test is most appropriate for this lesson objective :  
“The student will be able to identify the elements with their symbols from the periodic table.”
- A. Binary test
  - B. Short answer
  - C. Completion
  - D. Matching type

66. Study this test then answer the question:

**Directions:** Match the following in order to complete the sentences on the left.

- |   |                                      |
|---|--------------------------------------|
| ___1. Plato insisted that government was                      | A. The Prince.                       |
| ___2. Machiavelli wrote about achieving political unity       | B. desirable and inevitable          |
| ___3. Hobbes argued that human nature made absolute Monarchy  | C. a science requiring experts.      |
| ___4. Marx was a German philosopher and economist who founded | D. organized along industrial lines. |
|   | E. Communism.                        |

Which rule in test construction is violated by this test?

- A. Avoid grammatical or other clues to the correct response.
  - B. The premise should be in Column 2
  - C. The options should be more.
  - D. The options should be numbered.
67. The components of a multiple choice item are a
- A. stem and several foils
  - B. correct answer and several foils
  - C. stem, a correct answer and some foils
  - D. stem and a correct answer

68. Which are TRUE of portfolio assessment?

- I. Promotes student self-evaluation, reflection, and critical thinking
- II. Measures performance based on genuine samples of student work
- III. Provides flexibility in measuring how students accomplish their learning goals

IV. Enables teachers and students to share the responsibility for setting learning goals and for evaluating progress toward meeting those goals

- A. I and II
- B. B. I and IV
- C. II and III
- D. I, II, III, and IV

69. Which statements apply to portfolio?

- I. They clearly reflect stated learner outcomes identified in the core or essential curriculum that students are expected to study.
- II. They contain samples of work that stretch over an entire marking period, rather than single points in time.
- III. They contain works that represent a variety of different assessment tools.
- IV. They contain a variety of work samples and evaluations of that work by the student, peers, and teachers, possible even parents' reactions.

- A. I and II
- B. I and IV
- C. II and III
- D. I, II, III, and IV

70. To which type of assessment does assessment OF learning refer?

- A. Summative
- B. Diagnostic
- C. Formative
- D. Standardized

71. To which type of assessment does assessment FOR learning refer?

- A. Summative
- B. Diagnostic
- C. Formative
- D. Standardized

72. The phrase assessment AS learning refers to \_\_\_\_\_.

- A. Self-assessment
- B. Summative assessment
- C. Formative assessment
- D. Standardized assessment

73. It is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes. Which type of assessment is referred to ?

- A. Summative
- B. Formative
- C. Self-assessment
- D. Diagnostic

74. In what way are students involved in the formative process?

**Students:**

- A. monitor their ongoing progress.

- B. set their own learning goals.
  - C. show proof of learning.
  - D. evaluate instruction.
75. What does a portfolio include?
- I. Student participation in selecting contents
  - II. Criteria for selection
  - III. Criteria for judging merits
  - IV. Evidence of a student's self-reflection
- A. I, II, III and IV                      C. I, II and III  
 B. I and II                                  D. I only
76. Which type of portfolio includes the best and weakest of student work?
- A. Development portfolio              C. Showcase portfolio  
 B. Process portfolio                      D. Performance portfolio
77. Which type of portfolio emphasizes students' reflection on their learning, including the use of reflective journals, think logs, and related forms of metacognitive processing?
- A. Development portfolio              C. Showcase portfolio  
 B. Process portfolio                      D. Performance portfolio
78. Which type of portfolio is best used for summative evaluation of students' mastery of key curriculum outcomes?
- A. Development portfolio              C. Showcase portfolio  
 B. Process portfolio                      D. Performance portfolio
79. Which is NOT TRUE of an appropriate assessment method?
- A. Requires minimum effort to develop
  - B. Is a valid and reliable measure of the learning targets
  - C. Method matches the learning targets
  - D. Yields evidence that the learning targets were met
80. A teacher constructs a test blueprint prior to writing items for a test, to ensure test\_\_\_\_\_
- A. validity
  - B. scorability
  - C. reliability
  - D. fairness

81. Which of the following actions contributes to enhancing a test's validity?
- A. Comparing the scores of students with different demographic characteristics
  - B. Re-administering the test to the same examinees at least two weeks apart
  - C. Comparing scores on the same examinees before and after instruction
  - D. Having other people examine the items to make sure there is nothing offensive
82. Which of the following actions contributes to the improvement of test reliability?
- A. Increase the number of items and score points on the test.
  - B. Decrease the number of inferences or decisions based on the results.
  - C. Increase the number of objectives measured by the test.
  - D. Decrease the number of different types of item formats used.
83. The learning specialist administers a set of tests to identify specific weaknesses in Nick's learning processes. These types of tests are known as \_\_\_\_\_ tests.
- A. diagnostic
  - B. aptitude
  - C. intelligence
  - D. formative
84. Which is TRUE of authentic tests?
- A. Use norm-referenced grading
  - B. Emphasize breadth of knowledge more than depth
  - C. Maximize comparisons among students
  - D. Entail self-assessment
85. Which of these BEST defines assessment?
- A. Judging the quality of student performance
  - B. Specifying what students should know and be able to do
  - C. Reporting what students have learned or not learned
  - D. Gathering information about students to make decisions
86. Which of these BEST explains pre-instructional assessment?
- A. Establishes baselines grades for a unit of instruction
  - B. Monitors student learning during a unit of instruction
  - C. Determines what students learned as a result of instruction
  - D. Guides learning targets and activities for a unit of instruction

87. If Teacher Marian wants to determine what is considered acceptable evidence of mastery of the specific learning targets, it is best for her to
- plan a schedule when assessments will be administered
  - plan how to give students' feedback based on evidence
  - announce to students what the grading system will be
  - determine what assessment methods will yield the evidence**
88. Which is a feature of an imperfect matching type of test?
- An answer may be repeated.**
  - The items in the 2 columns are not equal.
  - Items on the left and right column have a one to one correspondence.
  - Its premises are vague.
89. Which of these is NOT a characteristic of recent trends in assessment?
- On-going gathering of information throughout instruction
  - Publicly-known criteria for evaluating performance
  - Variety of methods used to gather information about performance
  - Gathering performance information about isolated skills or facts**
90. Which are TRUE of portfolio assessment?
- Promotes student self-evaluation, reflection, and critical thinking
  - Measures performance based on genuine samples of student work
  - Provides flexibility in measuring how students accomplish their learning goals
  - Enables teachers and students to share the responsibility for setting learning goals and for evaluating progress toward meeting those goals
- I and II**
  - I and IV
  - II and III
  - I, II, III, and IV
91. Which is/are non-restricted essay test item/s?
- Should the lawmakers' pork barrel be abolished?
  - What is meant by pork barrel?
  - Trace the origin of pork barrel in the Philippines?
- I only**
  - II and III
  - II only
  - III only
92. Which of the following statements best defines content standard?
- When students should accomplish the content standards**

- B. The rate at which students should accomplish the content standards
  - C. How well students should accomplish the content standards
  - D. What students should know and be able to do
93. According to the current test theory, test validity is best described as
- A. the appropriateness of the inferences or use that results from the assessment
  - B. the equitableness with which the assessment measures members of different groups
  - C. the consistency with which the test measures something over time
  - D. the ease with which the assessment may be scored fairly by different people
94. “Validity is a unitary concept”. What does this expression mean?
- A. Different types of tests are brought together to form a single, valid assessment.
  - B. A valid test is one that has a single purpose.
  - C. There are not different kinds of validity, just different kinds of evidence of validity.
  - D. A valid test is one that has universal application in a variety of situations.
95. What define and differentiate levels of performance?
- |               |            |
|---------------|------------|
| A. Portfolios | C. Tests   |
| B. Researches | D. Rubrics |
96. Which are TRUE of K to 12 assessment of the Department of Education?
- I. It emphasizes formative learning.
  - II. It is standard-based.
  - III. It gives more weight to knowledge.
  - IV. It gives less weight to understanding.
- |               |               |
|---------------|---------------|
| A. I and IV   | C. III and IV |
| B. II and III | D. I and II   |

## 4.2 Interpret assessment results

152. One of the test items constructed by Teacher Gina was suspected of being biased. This happens when \_\_\_\_\_.
- A. students of equal ability with different demographic characteristics have an unequal probability of answering an item correctly
  - B. students with different demographic characteristics have an equal probability of answering an item correctly
  - C. students who have different levels of knowledge and skills have an equal probability of answering an item correctly

- D. students with equal exposure to the content being tested have an unequal probability of answering an item correctly
153. The statistical data showed that there was a significant increase in scores that resulted from the use of the new language arts program by the Language teachers. This result indicates that \_\_\_\_\_.
- A. the new program is the best available program
  - B. the new program is suitable for all students
  - C. the results did not happen by chance
  - D. the results have been replicated in other testing situations
154. Which best describes what test reliability refers to?
- A. Appropriateness for measuring the learning targets
  - B. Evidence of fairness to all examinees
  - C. Consistency of test results
  - D. Stability of scores over time
155. Dr. Legaspi aims to determine how well the students in his eighth grade class are reading as compared to other students in the division. As a teacher, Dr. Legaspi should examine the results of which of the following kinds of tests?
- A. A criterion-referenced reading test
  - B. A norm-referenced reading test
  - C. An aptitude test
  - D. An informal reading inventory
156. Which is the TRUEST representation of a mesokurtic distribution?
- A. Distribution where most scores fall in the middle and some scores fall higher and lower.
  - B. Normal distribution of student's scores
  - C. Distribution where some scores fall in the middle but most fall toward the extremes
  - D. There is no "normal" distribution when dealing with the social sciences.
97. Which test has a higher diagnostic value?
- A. Restricted essay
  - B. Multiple choice
  - C. Problem solving
  - D. Extended essay

98. Which test has a higher diagnostic value?
- A. Binary test
  - B. Restricted essay
  - C. Extended essay
  - D. Problem solving
99. If, before you proceed, you want to know what your students already know about your cognitive lesson for the day, which test may serve best your purpose?
- A. Objective test
  - B. Short answer test
  - C. Essay
  - D. Performance test
100. Which goes with assessment of learning?
- A. Given at the end of a grading period
  - B. Self –assessment
  - C. Guides teacher in adjusting instruction
  - D. Highly reflective
101. The following go with assessment for learning EXCEPT:
- A. Given at the end of a grading period
  - B. Given at the beginning
  - C. Guides teacher in adjusting instruction
  - D. Done in the process of instruction
102. Which is TRUE of diagnostic assessment?
- A. It has a formative purpose.
  - B. It as a summative purpose.
  - C. It is for grade purposes.
  - D. It is given at the end of a lesson.
103. Which is TRUE of diagnostic assessment?
- A. It as a summative purpose.
  - B. It is for grade purposes.
  - C. It is given at the end of a lesson.
  - D. It enables teachers to provide the scaffolding that students need.
104. Some experts believe that assessment has more influence on the way students study and learn than any other aspect of the teaching-learning process. What is an implication of this to instruction and assessment?
- A. Measure higher level outcomes to motivate students to engage in higher level learning.

- B. Avoid grading tests.
  - C. Make students compete against one another to meet high expectations.
  - D. Make use of descriptive grading system.
105. Which one brings down the reliability of test item?
- A. Ambiguous test item
  - B. Long test
  - C. Combining objective and essay test
  - D. Easy mere recall items
106. Which one brings down the reliability of test item?
- A. Long test
  - B. Combining objective and essay test
  - C. Unclear test direction
  - D. Pre recall test items
107. Which one brings down the reliability of test item?
- A. Long test
  - B. Combining objective and essay test
  - C. Pre recall test items
  - D. Inconsistent scoring
108. Which is a way of increasing the reliability of assessment?
- A. Use sufficient number of items.
  - B. Limit your number of test items to 10 or less.
  - C. Do away with test direction.
  - D. Think of the criteria of scoring an essay test after you have read the paper of the best student.
109. Which is a way of increasing the reliability of assessment?
- A. Limit your number of test items to 10 or less.
  - B. Specify scoring criteria in advance .
  - C. Do away with test direction.
  - D. Ask high level questions
110. Which is a way of increasing the reliability of assessment?
- A. Limit your number of test items to 10 or less.
  - B. Ask high level questions
  - C. Do away with test direction.

D. Score test without knowing the owner of the test that you are assessing.

111. Is an unreliable assessment valid?

- A. Yes.
- B. Depends on the difficulty index of the test item.
- C. No, because it gives inconsistent information even if it is aligned with assessment.
- D. Information is insufficient to give a categorical answer.

Study this matching type of test then answer the questions that follow:

- |   |   |
|---|---|
| 1. alliteration                                     | A. Author of "Macbeth"  |
| 2. hyperbole  | B. Author of "The Raven"                                      |
| 3. metaphor   | C. Understanding is like a light bulb coming on in your head. |
| 4. Shakespeare                                      | D. Name given to the Philippines                              |
| 5. personification made.                            | E. That's the most brilliant comment ever made.               |
| 6. simile of his                                    | F. His oratory was a bellow from the bowels of his soul.      |
| 7. Edgar Allan Poe advantageous                     | G. Appropriate attitudes are always advantageous              |
| 8. "Pearl of the Orient Seas"                       | H. His eyes are limpid pools of longing.                      |
| 9. Francisco Baltazar                               | I. He stood as straight as a rod.                             |
| 10. I 'll never get this stuff, no matter what I do |   |

112. What is a flaw of the matching type of test above?

- A. The items are homogeneous.
- B. The items are heterogeneous.
- C. The items are easy.
- D. The items are difficult.

113. What is a flaw of the matching type of test above?

- A. The items in both columns are not equal in number.
- B. The items in Column 1 are more in number than those in Column 2.
- C. The items are very few.

D. The items in both columns are in the wrong columns.

114. What is a flaw of the matching type of test above?

A. The items in both columns are not equal in number.

B. The items in Column 1 are more in number than those in Column 2.

C. The items are very few.

D. Some items can be answered by process of elimination.

115. Which should be done to improve on the flawed matching type of test above?

I. Make the items homogeneous.

II. make the items heterogeneous.

III. The items in Column 1 should be in Column 2 and the items in Column 2 should be in Column 1.

A. I and III

C. I and II

B. II and III

D. I, II and III

116. Which should be done to improve on the flawed matching type of test above?

I. Make the items homogeneous.

II. The items in Column 1 should be in Column 2 and the items in Column 2 should be in Column 1.

III. Alphabetize the items in Column 2 after they have been exchanged with those in Column 1.

A. I and III

C. I and II

B. II and III

D. I, II and III

117. To make the items homogeneous, which items should you retain ?

A. Those items on figures of speech

B. The names of people

C. The names of people and place

D. Those items on figures of speech and names of people

118. Is it necessary to make the items in both Columns equal?

A. No, if the matching test will be limited only to the given figures of speech and the sample statements.

B. Yes.

C. Yes, for every item to have an answer.

D. If you make a total of 15 test items.

119. What does the above matching type of test lack?



125. When you ask questions to lead students to discover or create new connections, generalization and patterns, or combine ideas to form a new whole you ask students to \_\_\_\_\_.
- A. analyze  
B. evaluate  
C. comprehend  
D. synthesize
126. If you ask questions that will make students identify the organization and patterns within a system by identifying its component parts you are asking questions on \_\_\_\_\_.
- A. synthesis  
B. application  
C. analysis  
D. evaluation
127. A question asking students to classify is a question on \_\_\_\_\_.
- A. synthesis  
B. application  
C. analysis  
D. evaluation
128. Sample Question Frames such as “Do you agree...?”, “How would you decide about...?”, “What priority would you give...?” belong to questions on \_\_\_\_\_.
- A. analysis  
B. application  
C. synthesis  
D. evaluation
129. If you want students to use evidence and reasoned argument to judge how well a proposal would accomplish a particular purpose and /or resolve controversies or differences, you ask \_\_\_\_\_ questions.
- A. analysis  
B. synthesis  
C. application  
D. evaluation
130. Multiple choice tests can be used to test the ability to:
- I. recall memorized data  
II. apply theory to everyday situation  
III. apply theory to novel situations  
IV. use judgment in analyzing, weighing and evaluating
- A. I only  
B. I and II only  
C. I, II and III only  
D. I, II, III and IV
131. Multiple choice items provide highly reliable test scores because \_\_\_\_\_.
- A. they do not place a high degree of dependence on the students reading ability

- B. they place a high degree of dependence on a teacher's writing ability
- C. they are a subjective measurement of student achievement
- D. they allow a wide sampling of content and a reduced guessing factor

Read this test item then answer the question:

You should:

- A. always decide on an answer before reading the alternatives
- B. always review your marked exams
- C. never change an answer
- D. always do the multiple choice items on an exam first

132. Why is the above multiple choice item undesirable? Because \_\_\_\_\_.
- A. a direct question is more desirable than an incomplete statement.
  - B. there is no explicit problem or information in the stem.
  - C. the alternatives are not all plausible
  - D. the options vary in length

133. How will you improve on this test item?

In national elections in the Philippines, the President is officially:

- A. chosen by the people.
- B. chosen by members of Congress.
- C. chosen by the House of Representatives.
- D. chosen by the Electoral College."

- A. Transfer "chosen by" to the introductory sentence in the stem.
- B. Add a 5th option.
- C. Make it higher level.
- D. Improve its phraseology.

134. Study these test items then answer the question.

**Undesirable:**

While ironing her formal attire, Jane burned her hand accidentally on the hot iron. This was due to a transfer of heat between...

**Desirable:**

Which of the following ways of heat transfer explains why Jane's hand was burned after she touched a hot iron?

Which guideline in test construction was observed ?

- A. Include in the stem any word(s) that might otherwise be repeated in each alternative.
- B. Present a definite, explicit and singular question or problem in the stem.
- C. Make all alternatives plausible and attractive to the less knowledgeable or skillful student.
- D. Eliminate excessive verbiage or irrelevant information from the stem.

135. Compare these two multiple choice test items:

***Undesirable test item***

The daily minimum required amount of milk that a 10 year old child should drink is

- A. 1-2 glasses.
- B. 2-3 glasses.
- C. 3-4 glasses.
- D. at least 4 glasses.

***Improved version***

What is the daily minimum required amount of milk a 10 year old child should drink?

- A. 1 glass.
- B. 2 glasses.
- C. 3 glasses.
- D. 4 glasses.

How was the test item improved?

- A. The alternatives were made mutually exclusive.
- B. The alternatives were made plausible and attractive.
- C. Eliminate excessive verbiage from the stem.
- D. Present a definite, explicit and singular question or problem in the stem.

136. Two ideas can be included in a true-false statement if the purpose is to show cause and effect.

- A. Yes

- B. No
- C. Depends on the thought of the statement
- D. Sparingly

137. Read this flawed T-F test than answer the question:

Water will boil at a higher temperature if the atmospheric pressure on its surface is increased and more heat is applied to the container.

Which guideline was not observed in the formulation of such item?

- A. Express a single idea in each test item.
- B. Base true-false items upon statements that are absolutely true or false, without exceptions .
- C. Avoid the use of extreme modifiers or qualifiers
- D. Avoid the use of unfamiliar vocabulary.

138. Which is an improved version of this T-F test item?

Water will boil at a higher temperature if the atmospheric pressure on its surface is increased and more heat is applied to the container.

- I. Water will boil at a higher temperature if the atmospheric pressure on its surface is increased.
- II. Water will boil at a higher temperature if more heat is applied to the container.
- III. If the atmospheric pressure on its surface is increased and more heat is applied to the container, water will boil at a higher temperature

- A. I only
- B. I and II
- C. III only
- D. I, II and III

139. Why is this test item undesirable?

According to some politicians, the raison d'etre for capital punishment is retribution.

Which guideline in the construction of T-F test was not observed?

- A. Arrange the statements so that there is no discernible pattern of answers.

- B. Avoid the use of extreme modifiers or qualifiers
- C. Avoid the use of unfamiliar vocabulary.
- D. Do not lift statements verbatim from textbooks.

140. Which guideline/s in T-F test construction is/are CORRECT?

- I. If a proposition expresses a relationship, such as cause and effect or premise and conclusion, present the correct part of the statement first and vary the truth or falsity of the second part.
  - II. Make true and false items of approximately equal average length throughout the test.
  - III. Randomize the sequence of true and false statements.
- A. I, II and III                      C. II only  
B. I and II                              D. II and III

141. When presenting words and their definitions in a matching test, which column should contain the definitions, which are longer than the words?

- A. Column B                              C. Column A  
B. Either Column A or B              D. Cannot be determined

142. In a matching test, in which column are items arranged in chronological order?

- A. Column B                              C. Column A  
B. Either Column A or B              D. Cannot be determined

143. In a matching test in which Column is the premise found?

- A. Column B                              C. Column A  
B. Either Column A or B              D. Cannot be determined

144. In a matching test, in which Column are items designated by numbers?

- A. Column B                              C. Column A  
C. Both Columns A and B              D. Cannot be determined

145. For the three learning objectives listed below, decide which one would be appropriate for a matching exercise as a method of assessment?

- A. The student will be able to evaluate a paragraph .
- B. The student will be able to differentiate between words that are spelled correctly and those spelled incorrectly.
- C. The student will be able to identify the elements with their symbols from the periodic table.
- D. A and B

146. Read this item then answer the question.

Put blanks at the \_\_\_\_\_ of the statement rather than the \_\_\_\_\_.

What is defective with this item?

- A. Overmutilated
- B. Insignificant
- C. Very short
- D. Not fit for completion type of test

147. Read the question then answer

Original version: "Every atom has a central \_\_\_\_\_  
called a nucleus".

Improved version : "Every atom has a central core called a(n)  
\_\_\_\_\_."

Which was the improvement ?

- A. The one omitted (the blank) was the significant word.
- B. Made shorter
- C. Made more direct
- D. Made shorter and more direct

148. Why should the blank in a completion test be near the end or at the end?

- A. Question first before answer
- B. Neater that way
- C. For greater focus
- D. To ensure clarity.

149. "What if... " questions are meant to develop students'

- A. analytical thinking
- B. synthetic thinking
- C. creative thinking
- D. comprehension

150. Assessment is integral to the teaching-learning process. What define and differentiate levels of performance?

- A. Portfolios
- B. Researches
- C. Tests
- D. Rubrics CLUSTER 4.1

151. The Department of Education mandates that the assessment process in evaluating learning is holistic. Which best describe the process?

- I. It emphasizes formative assessment
  - II. It is standard-based.
  - III. It gives more weight to knowledge.
  - IV. It gives less weight to understanding.
- A. I and IV
  - B. II and III
  - C. III and IV
  - D. I and II CLUSTER 4.1

91. Which is/are non-restricted essay test item/s?

- I. Should the lawmakers' pork barrel be abolished?
  - II. What is meant by pork barrel?
  - III. Trace the origin of pork barrel in the Philippines ?
- A. I only
  - B. II and III
  - C. II only
  - D. III only

92. Which of the following statements best defines content standard?

- A. When students should accomplish the content standards
- B. The rate at which students should accomplish the content standards
- C. How well students should accomplish the content standards
- D. What students should know and be able to do

93. According to the current test theory, test validity is best described as

- A. the appropriateness of the inferences or use that results from the assessment
- B. the equitableness with which the assessment measures members of different groups
- C. the consistency with which the test measures something over time
- D. the ease with which the assessment may be scored fairly by different people

94. "Validity is a unitary concept". What does this expression mean?

- A. Different types of tests are brought together to form a single, valid assessment.
- B. A valid test is one that has a single purpose.
- C. There are not different kinds of validity, just different kinds of evidence of validity.

- D. A valid test is one that has universal application in a variety of situations.
95. What define and differentiate levels of performance?  
A. Portfolios                      C. Tests  
B. Researches                      D. Rubrics
96. Which are TRUE of K to 12 assessment of the Department of Education?  
I. It emphasizes formative learning.  
II. It is standard-based.  
III. It gives more weight to knowledge.  
IV. It gives less weight to understanding.  
A. I and IV                      C. III and IV  
B. II and III                      D. I and II

## 4.2 Interpret assessment results

152. One of the test items constructed by Teacher Gina was suspected of being biased. This happens when \_\_\_\_\_.  
A. students of equal ability with different demographic characteristics have an unequal probability of answering an item correctly  
B. students with different demographic characteristics have an equal probability of answering an item correctly  
C. students who have different levels of knowledge and skills have an equal probability of answering an item correctly  
D. students with equal exposure to the content being tested have an unequal probability of answering an item correctly
153. The statistical data showed that there was a significant increase in scores that resulted from the use of the new language arts program by the Language teachers. This result indicates that \_\_\_\_\_.  
A. the new program is the best available program  
B. the new program is suitable for all students  
C. the results did not happen by chance  
D. the results have been replicated in other testing situations
154. Which best describes what test reliability refers to?  
A. Appropriateness for measuring the learning targets  
B. Evidence of fairness to all examinees  
C. Consistency of test results  
D. Stability of scores over time

155. Dr. Legaspi aims to determine how well the students in his eighth grade class are reading as compared to other students in the division. As a teacher, Dr. Legaspi should examine the results of which of the following kinds of tests?

- A. A criterion-referenced reading test
- B. A norm-referenced reading test**
- C. An aptitude test
- D. An informal reading inventory

156. Which is the **TRUEST** representation of a **mesokurtic** distribution?

- A. Distribution where most scores fall in the middle and some scores fall higher and lower.
- B. Normal distribution of student's scores**
- C. Distribution where some scores fall in the middle but most fall toward the extremes

D. There is no "normal" distribution when dealing with the social sciences.

157. Study the score distribution below then answer the following questions:

Student	1	2	3	4	5	6	7	8	9
Score	15	18	16	14	15	15	12	17	90

What is the median?

- A. 15**
- B. 14
- C. 17
- D. 16

158. What is the mean?

- A. 23.55**
- B. 26
- C. 24
- D. 25

159. Is there a mode?

- A. Yes.
- B. Yes, there are 3.
- C. Yes, only one, 15.**
- D. Yes, only one, 12.

160. In this case, which is/are a more appropriate measure/s of central tendency?

- A. Mean
- B. Mode
- C. Median
- D. Mode and median**

161. Which is TRUE of the score distribution?

- A. The range is quite high.**

- B. Standard deviation is zero.
- C. It is bimodal.
- D. It is trimodal.

162. If the criterion of success is a score of 75, what can you say of the given score distribution?

- A. Only one met the criterion of success.
- B. One is below the criterion of success.
- C. Not one met the criterion of success.
- D. Only eight met the criterion of success.

163. The Figure below shows a comparison of the tutorial group sizes you had for the 1st and 2nd semesters. Study them then answer the questions.

With drawing

Can it be said that the size of the tutorial group is more varied or dispersed in the 2nd semester?

- A. Yes.
- B. No.
- C. No, as the Figures show.
- D. Need more data to be able to answer

164. Can it be said that the size of the tutorial group is less varied or dispersed in the 1st semester?

- A. Yes.
- B. No.
- C. No, as the Figures show.
- D. Need more data to be able to answer

165. What is the range of the size of the tutorial group for the 1st sem?

- A. 3
- B. 5
- C. 4
- D. 2

166. What is the range of the size of the tutorial group for the 2nd sem?

- C. 3
- D. 5
- C. 4
- D. 2

167. Which statement is TRUE of range?

- A. It tells how spread a set of data is.
- B. It tells how concentrated a set of data is.
- C. It is computed by deducting the highest number from the lowest.
- D. It is the most sophisticated way of computing for the variability of data.

Study the Table below then answer the ff. questions:

With table

168. You are interested to compute the variability of the students' marks in both coursework and examination. What are you interested in?

- A. How the scores converge around the mean
- B. How spread the scores are
- C. How close the coursework marks are to the examination marks
- D. How far the coursework marks are to the examination marks

169. What is/are TRUE of the score distribution?

- I. The highest coursework mark is 48 and the lowest is 27 giving a range of 21.
  - II. In the examination, the highest mark is 47 and the lowest is 12 producing a range of 35.
  - III. There is a wider variation in the students' performance in the examination than in the coursework for this module.
- A. I only
  - B. II only
  - C. II and III
  - D. I, II and III

170. Suppose you add zero as score in both coursework and examination marks, will the range be still typical of the marks?

- A. No more
- B. Can't be determined
- C. Very much
- D. Yes only for coursework marks



- B. Score is the cut off itself.
- C. Score is below the cut off score.
- D. Score is very high.

176. Study the Figures below then answer the questions.

With figures

Which is TRUE of the score distributions shown by the figures?

- A. The Figure on the right has a higher SD.
- B. The Figure on the left has a higher SD.
- C. The score distributions are not symmetrical.
- D. The scores of the Figure on the left are more spread.

177. Which is TRUE of the score distributions shown by the figures?

- A. The Figure on the left shows leptokurtic score distribution.
- B. The Figure on the right shows leptokurtic score distribution.
- C. The score distributions are not symmetrical.
- D. The scores of the Figure on the right are less spread.

178. Which is TRUE of the score distributions shown by the figures?

- A. The Figure on the right comes close to a platykurtic score distribution.
- B. The Figure on the right platykurtic score distribution.
- C. The score distributions are not symmetrical.
- D. The scores of the Figure on the left are more spread.

179. The scores are in the middle in a \_\_\_\_\_ score distribution.

- A. normal
- B. mesokurtic
- C. leptokurtic
- D. platykurtic

180. If the scores have a large spread without a lot of scores in the middle, we have a \_\_\_\_\_ score distribution.

- A. normal
- B. mesokurtic
- C. leptokurtic
- D. platykurtic

181. Study the Figures then answer the questions:

Which is TRUE of the positively skewed distribution?

- A. Most scores are high.
- B. Most scores are low.**
- C. Test may be easy.
- D. Students learned very much.

182. Which is TRUE of the negatively skewed distribution?

- A. Most scores are high.**
- B. Most scores are low.
- C. Test may be difficult.
- D. Students learned very little.

183. If score distribution follows a normal curve, which apply?

- I. Unimodal (one mode)
  - II. Symmetrical -just as many people above the mean as below the mean
  - III. Bell shaped
  - IV. Mean, Mode, and Median are all located in the center
- A. I, II, III and IV**                      C. I and II  
B. I and III                                  D. I, II

184. When should you not use the mean to determine central tendency? When \_\_\_\_.

- I. there are values that are unusual compared to the rest of the data set by being especially small or large in numerical value
  - II. there are few scores.
  - III. when score distribution is skewed
- A. I only**                                      **C. I and III**  
B. II only                                      D. I, II and III

185. You got a 76 (of 100) and your friend got 82 (also of 100). Did your friend do better than you?

- I. Not necessarily
  - II. Yes
  - III. Other factors such as SD, level of difficulty need to be considered.
- A. I and III**                                      C. II only  
B. I only                                        D. III only

186. Suppose that in a 100-item test, your score is 76 and in another 100-item test your friend got a score of 82? Did your friend do better if the mean mark in your class

was 54 and the standard deviation was 20 and the mean mark in your friend's class is 72 and the standard deviation was 15. Your Z score is  $(76-54)/20 = 1.1$ . Your friend's Z score is  $(82-72)/15 = 0.67$ .

- A. You did better than your friend because your mark was more standard deviations above the class mean.
- B. Your friend did better because his raw score (82) is higher than yours (76).
- C. It cannot be told.
- D. More data like IQ are needed on which to base answer.

187. Who performed better – Student A who graduated 37th out of a class of 250 or Student B who graduated 12th in a class of 60?

- A. Student A
- B. Cannot be determined
- C. Student B
- D. The two performed equally

188. What does 85th percentile mean of an examinee?

- A. S/he belongs to the upper 15% of the examinees.
- B. S/he belongs to the upper 85% of the examinees.
- C. S/he got a raw score of 15 out of an 85-item test.
- D. S/he got a raw score of 85 out of a 100-item test.

189. Study the Table then answer the questions that follow.

With tables

Which are TRUE of the data?

- I. About two-thirds of the upper group but only one-third of the lower group answered the item correctly.
  - II. The examinees who answered the item incorrectly did not select the incorrect response in equal proportions.
  - III. The examinees concentrated on a single incorrect option
  - IV. Option two seems to be the most attractive incorrect option, especially to the upper and middle groups.
- A. I, II and III
  - B. I and II
  - C. I, II, III and IV
  - D. II and III

190. Which option needs to be more attractive?

- A. Option 4
- B. Option 2
- C. Option 1
- D. Option 3

191. Which is/are TRUE of discrimination index? The index of discrimination ....

- I. is useful in a test is to produce a spread of scores
- II. is not useful in mastery model tests.
- III. is useful in criterion-referenced tests.

- A. I and II
- B. II only
- C. I only
- D. I, II and III

192. Study the tabular data then answer the questions: Question #1 and #2.

With table

Which are TRUE of the data:

- I. Option A of Question #1 was not a very good distracter .
- II. The difficulty index of item 1 is .80.
- III. The difficulty index of item 2 is .40).
- IV. Distractor B of item 2 should be carefully studied by the item writer.

- A. I and II
- B. I and III
- C. II and III
- D. I, II, III and IV

193. A rough “rule-of-thumb” is that if the item difficulty is more than .75, it is an easy item; if the difficulty is below .25, it is a difficult item. How do you regard item 1?

- A. Moderately easy
- B. Easy
- C. Extremely easy
- D. Cannot be classified

194. Study the tabular data then answer the following questions. .

Item	# Correct (Upper group)	# Correct (Lower group)	Difficulty (p)	Discrimination (D)
Question 1	4	4	.80	0
Question 2	0	3	.30	-0.6
Question 3	5	1	.60	0.8

Which is the best question?

- A. Question # 3
- B. Question # 2
- C. Question # 1
- D. None

195. Which is the most difficult test item?

- A. Question # 2
- B. Question # 3
- C. Question # 1
- D. None

196. Which item should be changed or discarded?

- A. Question # 2
- B. Question # 3
- C. Question # 1
- D. None

Refer to the Table below to answer Items 1 and 2.

### Reading Comprehension Test Results

Section	English (MPS)	Filipino (MPS)
A	76	88
B	61	32
C	30	65
D	82	90
E	50	73

197. Which section/s should be given remedial instruction in English and Filipino?

- A. Sections B & C
- B. Sections C & E
- C. Section C only
- D. Section B only **CLUSTER 4.2**

198. Who among the teacher/s should be advised to conduct an action research?

- A. English teacher of Section C Filipino Teacher of Section B
- B. English teacher of Section E Filipino Teacher of Section C
- C. English Teacher of Section B only Filipino Teacher of Section C only

**CLUSTER 4.2**

**4.3 Utilize processed data and assessment results in reporting learners' performance to improve teaching and learning**

199. John tells his mother he got a score of 35 in a national achievement test. What other information is needed for that score to have meaning?
- I. The highest possible score
  - II. How the rest of the examinees compare with John in terms of score
  - III. No additional information needed
- A. III only
  - B. II only
  - C. I only
  - D. I and II
200. The range of a score distribution where 50 is the highest possible score is 50. What does this imply?
- A. The lowest score is zero.
  - B. The lowest score is 10.
  - C. The highest score is 100.
  - D. The midpoint is 50.
201. The range of a score distribution where 50 is the highest possible score is 50. What does this imply?
- A. The lowest score is 10.
  - B. The lowest score is 20.
  - C. The highest score is 50.
  - D. The midpoint is above 50.
202. Below are sound advice on how teachers should utilize assessment results, EXCEPT:
- A. feedback to the pupil should be about particular qualities of his work
  - B. give specific advice on what the pupil can do to improve
  - C. Compare pupil's score with those of the other pupils
  - D. compare scores to the mean score
203. Which is/are way/s by which a school may use assessment results?



208. Eighty percent (80%) of your students got scores 2SDs above the mean. What will you do?

- I. Find out how the remaining 20% performed.
- II. Focus on the 80% who performed well since they are the majority.
- III. Help the 20% for better performance

- A. I and III
- B. I and II
- C. I only
- D. I, II and III

209. After studying multiple assessment results, you discover that student B performed very poorly on all tests – written and unwritten. What should you do?

- I. Determine the cause of poor performance.
- II. Talk to her parents.
- III. Determine if she is a child with special needs.

- A. I, II and III
- B. I and II
- C. I and III
- D. II and III

#### 4.4 Demonstrate skills in the use of techniques and tools in assessing affective learning

210. Teacher D asks her students to write a summary of their most intimate thoughts about a film viewed to determine how much the film has changed their thoughts about racial discrimination. Which assessment tool does she employ?

- A. Journal entry
- B. Observation sheet
- C. Reflection worksheet
- D. Theme writing

211. Which is the best way to determine if students have imbibed punctuality after a lesson on punctuality?

- A. Reflection worksheet
- B. Journal entry
- C. Observation
- D. Debate on use of time

212. Teacher F wants to know how health conscious her students have become after a lesson on Health Tips. Which assessment tools is/are most appropriate?

- I. Checklist on health habits
  - II. Quiz on health tips
  - III. Observation
  - IV. Role plays
- A. I and II
  - B. I and III
  - C. II and III
  - D. I and IV

213. Which is the best assessment activity to test students' stand on same sex marriage?

- A. Debate
- B. Journal entry
- C. Observation
- D. Activity sheet

214. Which one may Teacher G use if she wants to measure students' opinions, attitudes and values on a psychometrically controlled scale.

- A. Semantic differential
- B. Journal entry
- C. Bipolar adjectives
- D. Interview

215. Which measures students' reactions to stimulus words and concepts in terms of ratings on bipolar scales defined with contrasting adjectives at each end?

- A. Semantic differential
- B. Bipolar scale
- C. Likert scale
- D. Journal

216. Which assessment tool for affective learning is given below?

- A. Journal
- B. Bipolar scale
- C. Likert scale
- D. Semantic differential

217. Which assessment tool offers opportunities to teachers to document students' growth over time?

- A. Anecdotal record
- B. Developmental record
- C. Bibliography
- D. Autobiography

218. Which is an on-the-spot recording that provides an inexpensive, non-threatening, assessable method of gathering a range of information about learners' academic, social, and emotional development?

- A. Anecdotal record
- B. Developmental record
- C. Bibliography
- D. Autobiography

219. Internalization of values can be assessed by the students ability to

- I. take a stand
- II. suggest appropriate actions
- III. observe the action.

- A. I only
- B. I and II
- C. II only
- D. I, II and III

220. With such knowledge and understanding, which of these best explains the limitations of assessing affective targets?

- A. There is no agreement on what affective targets are or what methods are appropriate for measuring them.
- B. The public (parents, media, etc.) has strong preferences for which affective targets should be measured and how they should be measured.
- C. There is no research to support the idea that affect can be quantified or measured.
- D. Methods for assessing these targets contain many sources of error, which leads to problems with reliability and validity.

221. Teacher Derek wanted to create an affective assessment. Which of these additional steps is most critical for Teacher Derek to include when creating an affective assessment?

- A. Record the date and the time the assessment information is collected.
- B. Have students respond anonymously.
- C. Use more than one measure and collect data at different points in time.
- D. Have another person conduct the assessment to avoid observer bias.

222. Which of these is NOT a recommended strategy for Mr. Wilson when he plans constructing questionnaires to assess affective targets?

- A. Avoid the use of extreme or absolute words or phrases
- B. Include many questions so students take the assessment seriously
- C. Ask one question at a time and avoid double-barreled questions
- D. Use present tense and avoid the use of negatives

223. According to the current test theory, test validity is best described as

- A. the appropriateness of the inferences or use that results from the assessment
- B. the equitableness with which the assessment measures members of different groups
- C. the consistency with which the test measures something over time
- D. the ease with which the assessment may be scored fairly by different people

224. Authors of textbooks recommend the use of multiple measures for assessing affective targets because \_\_\_\_\_.

- A. students' affects are unpredictable and constantly changing
- B. evidence from multiple measures counter the unreliability of a single measure

- C. bias seriously affects the interpretations made from multiple measures  
 D. individual and group responses or ratings may differ greatly
225. Teacher D asks her students to write a summary of their most intimate thoughts about a film viewed to determine how much the film has changed their thoughts about racial discrimination. Which assessment tool does she employ?  
 A. Journal entry  
 B. Observation sheet  
 C. Reflection worksheet  
 D. Theme writing
226. Which is the best way to determine if students have imbibed punctuality after a lesson on punctuality?  
 A. Reflection worksheet  
 B. Journal entry  
 C. Observation  
 D. Debate on use of time
227. Teacher F wants to know how health conscious her students have become after a lesson on Health Tips. Which assessment tools is/are most appropriate?  
 I. Checklist on health habits  
 II. Quiz on health tips  
 III. Observation  
 IV. Role plays  
 A. I and II  
 B. I and III  
 C. II and III  
 D. I and IV
228. Which is the best assessment activity to test students' stand on same sex marriage?  
 A. Debate  
 B. Journal entry  
 C. Observation  
 D. Activity sheet
229. Which one may Teacher G use if she wants to measure students' opinions, attitudes and values on a psychometrically controlled scale.  
 A. Semantic differential  
 B. Journal entry  
 C. Bipolar adjectives  
 D. Interview
230. Which measures students' reactions to stimulus words and concepts in terms of ratings on bipolar scales defined with contrasting adjectives at each end?  
 A. Semantic differential  
 B. Bipolar scale  
 C. Likert scale  
 D. Journal

#### 4.5 Assign students' marks / ratings

231. Principal Jeruta should NOT recommend which of the following practices to his teachers, to establish a fair and effective grading system?

- A. Give extra credit for completing nonacademic tasks.
- B. Base grades on student performance in relation to the targets.
- C. Use a sufficient number of assessments.
- D. Inform parents and students of grading procedures.

232. Mr. Jacob's grading is based on the normal curve, and thus, student performance is compared to the performance of others. What kind of grading is he using if he assigns quarter grades based on the normal bell curve?

- A. Criterion-referenced grading
- B. Informal assessment
- C. Authentic assessment
- D. Norm-referenced grading

233. Students' grades in School A is a function of the following: knowledge-15%; process or skills- 25%; understanding-30% and product and performance , 30%. What can you infer from the School A's grading system?

- A. Greatest emphasis is on direct evidence of learning.
- B. Recall is considered unnecessary.
- C. Understanding is enough proof of learning.
- D. Process is unnecessary.

234. Students' grades in School A is a function of the following: knowledge-15%; process- 25%; understanding-30% and product and performance , 30%. What can be inferred from School A's grading system?

- A. Grade is hard earned by every student.
- B. Grades aren't important.
- C. Grade is a function of multiple sources.
- D. Failing a student is de-motivating.

235. Here is School B's descriptive grading system.

Level of Proficiency	Numerical Grade
Beginning	74% and below
Developing	75-79%
Advancing in Proficiency	80-84%
Proficient	85-89%
Advanced	90% and above

What can be inferred from School A's grading system?

- A. There is no computing on the part of the teacher.
- B. This approves of mass promotion.
- C. All students are expected to advance in proficiency at the same pace.
- D. School is not content with numerical grades.

236. Study School B's descriptive grading system then answer the 3 questions that follow:

Level of Proficiency	Numerical Grade
Beginning (B)	74% and below
Developing (D)	75-79%
Advancing in Proficiency (AP)	80-84%
Proficient (P)	85-89%
Advanced (A)	90% and above

Under which level of proficiency is a student who struggles with his/her understanding and whose fundamental knowledge and/or skills has not been acquired or developed?

- A. Developing
- B. Approaching Proficiency
- C. Beginning
- D. Between Developing and Approaching Proficiency

237. Under which level of proficiency is a student who exceeds the core requirements in terms of knowledge, skills and attitudes and can transfer them automatically and flexibly through authentic tasks?

- A. Developing
- B. Approaching Proficiency
- C. Beginning
- D. Advanced

238. Under which level of proficiency is a student who possesses the minimum knowledge, skills and core understandings but needs help throughout the performance of his/her tasks?

- A. Developing
- B. Approaching Proficiency
- C. Beginning
- D. Advanced



3. Avoid complex or awkward word arrangements. Also, avoid use of negatives in the stem as this may add unnecessary comprehension difficulties.

**Example:**

**(Poor)**

Which of the following groups of citizens of the Philippines are not allowed to cast their votes in elections.

4. Avoid double negatives. Do not use with a negative stem since it becomes a double-negative

**Example:**

**(Poor):**

Which of the following is **least unlikely** to be a speculative purchase?

**(Better)**

Which of the following is likely to be the most speculative purchase?

Do not build upon other questions. keep questions independent of one another.

Do not supply answers to other questions. Avoid providing cues from one question to another;

Avoid negative stems. If negatives are necessary, they are emphasized with underlined, **bolded**, CAPITALIZED, *italicized*, and/or **colored indicators**. Word the stem positively; avoid negative phrasing such as “not” or “except.” If this cannot be avoided, the negative words should always be highlighted by underlining or capitalization: Which of the following is NOT an example .....

Do not introduce unfamiliar vocabulary and concepts in the test unless there is a relevant stated purpose in the test directions

Use the number of alternatives appropriate to a test item throughout the test, generally three to five (no necessity to use a consistent number throughout the test)

Sequence alternatives in logical or numerical order; Should there be no order, randomly assign correct answers in the sequence

Pay attention to grammatical consistency of all alternatives. Avoid giving clues through the use of faulty grammatical construction.

Vary position in sequence of alternatives. The answer must not only be A or B, or C or D.

Include common misconceptions as distracters

Include plausible content or viable cues in each distracter. Re-use key words from the correct alternative to make distracters more viable.

Avoid “All of the above”.

Do not use “None of the above” in a “best answer” question

Avoid giving unintended cues – such as making the correct answer longer in length than the distractor.

Test for important or significant information.

Focus on a single problem or idea for each exam question.

Keep the vocabulary consistent with the students’ level of understanding.

Avoid questions based on opinions.

Use multiple-choice to measure higher level thinking.

Be sensitive to cultural and gender issues.

State stem in either question form (When did World War II begin?) or the completion form (World War II began in \_\_\_\_\_.)

Avoid window dressing (excessive verbiage) in the stem.

Include the central idea and most of the phrasing in the stem.

Avoid giving clues such as linking the stem to the answer (... Is an example of *an*: test-wise students will know the correct answer should start with a vowel)

Place options in logical or numerical order.

Use letters in front of options rather than numbers; numerical answers in numbered questions may be confusing to students.

Keep options independent; options should not be overlapping.

Keep all options homogeneous in content.

Keep the length of options fairly consistent (preferably short)

Avoid specific determinates, such as *never* and *always*.

Make sure that there is one and only one correct option.

Use plausible distracters.

Incorporate common errors of students in distracters.

Avoid technically phrased distracters.

Use familiar yet incorrect phrases as distracters.

Avoid the use of humor when developing options.

## **Guidelines for Constructing True-False Test Items**

### **True-False**

1. Avoid trick statements with some minor misleading word or spelling anomaly. An experienced exam taker may profit from his detection of such a ploy, despite a lack of knowledge in the subject matter.

Example: Edgar Allen Poe wrote *The Raven*.

2. Use simple, direct language in declarative sentences.
3. Present the correct part of the statement first, and vary the truth or falsity of the second part if the statement expresses a relationship (cause, effect--if, then).
4. Statements must be absolute without qualification, subject to the true/false dichotomy without exceptions.
5. Every part of a true sentence must be “true” If any one part of the sentence is false, the whole sentence is false despite many other true statements.
6. Do not copy verbatim from textbooks.
7. Include background, qualifications, and context as necessary: “According to..., ....”
8. Avoid:

unfamiliar vocabulary and concepts

long strings of statements

ambiguous statements and generalizations that are open to interpretation

indefinite or subjective terms that are open to interpretation “a very large part” “a long time ago” “most” negative words and phrases: they can be confusing IF negatives are necessary, they are emphasized with underlined, bolded, CAPITALIZED, *italicized*, and/or colored indicators e.g.: “no” “not” “cannot” Drop the negative and read what remains to test your item

absolute words restrict possibilities. These imply the statement must be true 100% of the time and usually cue a “false” answer e.g.: “No” “never” “none” “always” “every” “entirely” “only”

negatives, qualifiers, absolutes, and long strings of statements

9. Base questions upon introductory material, as graphs, images, descriptions, problems, mediated objects, etc. to enhance assessment value, accommodate and empower those with alternative learning styles and evoke higher level thinking, analysis, or problem solving.
10. Write out essential content statements.
11. Make true and false statements equal in length.
12. Group questions by content
13. Build up to difficulty. Encourage with simpler questions first.
14. Randomize sequences of T/F responses. Avoid a discernable pattern
15. Vary the quantity of true/false statements from test to test recognizing that “true” is marked more often in guessing, and that assessing false statements tends to be more challenging.

### **Guidelines in Constructing Matching Type of Test**

For Perfect Matching type - Where an answer may not be repeated

Items for matching are homogeneous. If you match authors with their literary work just that. Don't mix provinces and their capitals nor dates and events.

The premises (questions which are longer than the answers) should be in Column 1 numbered 1, 2 3 and so on.

The options (answers) should be in Column 2 marked with the letters of the alphabet.

The options should be arranged alphabetically.

If options consist of numbers they should be arranged chronologically.

The number of options must be more than the number of premises so that answers cannot be arrived at by process of elimination.

Tell the examinees what to do by giving clear and complete direction.

Label Column 1 and Column 2.

Place options on the same page. Students should not have to turn back and forth for the answer.

### **For Imperfect Matching Type** – Where an answer or answers may be repeated

1. Caution the examinees by informing them that an answer or answers may be repeated. Include that in the direction.
2. Rules 1-7 for the perfect matching type also apply to imperfect matching type of test.

### **For Completion type**

1. The blank in completion questions should always be at the end of the stem.
2. There should only be one acceptable answer.
3. Ask questions on significant items. Avoid trivia. Be sure that the missing segment of the incomplete item is important.
4. Avoid over mutilated statements where there are too many blanks that there is no cue left. Do not omit too many words or the statement becomes unclear and will force the students to guess the answer.
5. Word the test item clearly and comprehensively enough to allow a student to answer correctly.
6. Use a direct question to test for comprehension of technical terms or knowledge of definitions.
7. Avoid using statements directly taken from the book.
8. Avoid giving grammatical clues to the correct answer.
9. Provide sufficient space on the answer sheet. But don't suggest the answer by the length of your blanks. Make the length of the blanks uniform. Provide blanks for the answers before the number.

### **For Essay**

#### **Types of Essay**

## **Restricted-response essays**

1. For learning outcomes not readily assessed objectively
2. Compared to extended-response questions, they target narrower learning outcomes, such as more specific mental processes (e.g., draws valid conclusions).

*Examples:*

How are asteroids and comets different?

What is recycling and what are items that can be recycled?

*Examples of Restricted Response*

- Describe two situations that demonstrate the application of the law of supply and demand. Do not use those examples discussed in class.
- State the main differences between the Vietnam War and previous wars in which the United States has participated.
- Why is the barometer one of the most useful instruments for forecasting weather? Answer in a brief paragraph. There are expected fixed answers . Not any answer is accepted.

## **Non-restricted Extended-Response Essay Questions**

- Freedom of response allows student to:
  - select information that are pertinent
  - organize answer in accordance with their best judgment
  - integrate and evaluate ideas as they deem appropriate.
- Place value on higher-order thinking skills.

*Examples:*

- If you were Charlotte, what would you have written in your web? Why?
- Is Templeton the rat a hero or a villain in the story? Why?
- Imagine that you and friend found a magic wand. Write a story about an adventure that you and your friend had with that magic wand.
- Compare developments in international relations in the administrations of President William Clinton and President George W. Bush. Cite examples when possible.

Place options on the same page. Students should not have to turn back and forth for the answer.

For Imperfect Matching Type – Where an answer or answers may be repeated

Caution the examinees by informing them that an answer or answers may be repeated.

Rules 1-7 for the perfect matching type also apply to imperfect matching type of test.

**For Completion type**

1. The blank in completion questions should always be at the end of the stem.
2. There should only be one acceptable answer.
3. Ask questions on significant items. Avoid trivia. Be sure that the missing segment of the incomplete item is important.

4. Avoid over mutilated statements where there are too many blanks that there is no cue left. Do not omit too many words or the statement become unclear and will force the students to guess the answer.
5. Word the test item clearly and comprehensively enough to allow a student to answer correctly.
6. Use a direct question to test for comprehension of technical terms or knowledge of definitions.
7. Avoid using statements directly taken from the book.
8. Avoid giving grammatical cues to the correct answer.
9. Provide sufficient space on the answer sheet. But don't suggest the answer by the length of your blanks. Make the length of the blanks uniform.

### **For essay**

#### **Type of Essay**

#### **Restricted - Restricted-response essays**

For learning outcomes not readily assessed objectively

Compared to extended-response questions, they target narrower learning outcomes, such as more specific mental processes (e.g., draws valid conclusions)

#### Examples:

How are asteroids and comets different?

What is recycling and what are items that can be recycled?

Examples of Restricted Response

- Describe two situations that demonstrate the application of the law of supply and demand. Do not use those examples discussed in class.
- State the main differences between the Vietnam War and previous wars in which the United States has participated.
- Why is the barometer one of the most useful instruments for forecasting weather? Answer in a brief paragraph. There are expected fixed answers. Not any answer is accepted.

**Non-restricted /Extended – These are open-ended in the sense that**

Extended-Response	Essay	Questions
<ul style="list-style-type: none"> <li>• Freedom of response allows student to select information that they think is pertinent, to organize the answer in accordance with their best judgment, and to integrate and evaluate ideas as they deem appropriate.</li> </ul>		
<ul style="list-style-type: none"> <li>• Places value on higher-order thinking skills.</li> </ul>		

Examples:

If you were Charlotte, what would you have written in your web? Why?

Is Templeton the rat a hero or a villain in the story? Why?

- Imagine that you and friend found a magic wand. Write a story about an adventure that you and your friend had with that magic wand.
- Compare developments in international relations in the administrations of President William Clinton and President George W. Bush. Cite examples when possible.

**Guidelines in Constructing Essay Tests**

**1. Clearly define the intended learning outcome to be assessed by the item.**

**2. Use essay questions to measure complex learning outcomes only.** The essay test should be reserved for those situations that require the student to *give* reasons, *explain* relationships, *describe* data, *formulate* conclusions, or in some other way produce the appropriate answer. Where supplying the answer is vital, a properly constructed restricted-response question is likely to be most appropriate. At the synthesis and evaluation levels of learning, the essay answer provides unique information.

**3. Relate the questions as directly as possible to the learning outcomes being measured.**

4. Indicate to the students the criteria (the rubric) to be used in evaluating the answer. For example, a parenthetical statement such as the following might be added: “Your answer will be evaluated in terms of its comprehensiveness, the relevance of its arguments, the appropriateness of its examples, and the skill with which it is organized.” This clarifies the task to the students without limiting their freedom, and makes the item easier to relate to clearly defined learning outcomes.

**5. Formulate questions that present a clear task to be performed.** Complex achievement is most apt to be called forth by such words as “why,” “describe,” “explain,” “compare,” “relate,” “contrast,” “interpret,” “analyze,” “criticize,” and “evaluate.”

**6. Provide ample time for answering and suggest a time limit on each question.** Time must be allowed for thinking as well as for writing and generous time limits should be provided. It may also be desirable to indicate page or word limits. Anything that will clarify the form and scope of the task without interfering with the intended outcomes is likely to contribute to more effective measurement.

[Modified from Gronlund, N.E. (2000). *Assessment of Student Achievement, 7th Ed.* Massachusetts: Allyn and Bacon.]

7. Give adequate directions as to the content of the desired response, i.e., don't just say "discuss," say "discuss in terms of x, y, and z."

Indicate the length of the response desired.

Don't provide optional questions, i.e., answer two of the following four questions. This results in different tests (if you do offer choices, make sure you are satisfied with different tests).

**State the criteria for grading**

**Present a reasonable task to students.**

### **Guidelines in Constructing Essay Tests**

1. Clearly define the intended learning outcome to be assessed by the item.
2. Use essay questions to measure complex learning outcomes only. The essay test should be reserved for those situations that require the student to *give* reasons, *explain* relationships, *describe* data, *formulate* conclusions, or in some other way produce the appropriate answer. Where supplying the answer is vital, a properly constructed restricted-response question is likely to be most appropriate. At the synthesis and evaluation levels of learning, the essay answer provides unique information.
3. Relate the questions as directly as possible to the learning outcomes being measured.
4. Indicate to the students the criteria (the rubric) to be used in evaluating the answer. For example, a parenthetical statement such as the following might be added: "Your answer will be evaluated in terms of its comprehensiveness, the relevance of its arguments, the appropriateness of its examples, and the skill with which it is organized." This clarifies the task to the students without limiting their freedom, and makes the item easier to relate to clearly defined learning outcomes.
5. Formulate questions that present a clear task to be performed. Complex achievements most apt to be called forth by such words as "why," "describe," "explain," "compare," "relate," "contrast," "interpret," "analyze," "criticize," and "evaluate."

6. Provide ample time for answering and suggest a time limit on each questions. Time must be allowed for thinking as well as for writing and generous time limits should be provided. It may also be desirable to indicate page or word limits. Anything that will clarify the form and scope of the task without interfering with the intended outcomes is likely to contribute to more effective measurement.
7. Give adequate directions as to the content of the desired response, i.e., don't just say "discuss," say "discuss in terms of x, y, and z."

Indicate the length of the response desired.

9. Don't provide optional questions, i.e., answer two of the following four questions. This results in different tests (if you do offer choices, make sure you are satisfied with different tests).
10. State the criteria for grading.
11. Present a reasonable task to students.

[Modified from Gronlund, N.E. (2000). *Assessment of Student Achievement, 7th Ed.* Massachusetts: Allyn and Bacon.]

## **Authentic forms of assessment**

A form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills -- Jon Mueller

"...Engaging and worthy problems or questions of importance, in which students must use knowledge to fashion performances effectively and creatively. The tasks are either replicas of or analogous to the kinds of problems faced by adult citizens and consumers or professionals in the field." -- Grant Wiggins -- (Wiggins, 1993).

### Traditional and Authentic Assessments Compared

<b>Traditional</b>	<b>Authentic</b>
Selecting a Response	Performing a Task
Contrived	Real-life
Recall/Recognition	Construction/Application

Teacher-structured -----Student-structured  
Indirect Evidence -----Direct Evidence

**Selecting a Response to Performing a Task:** On traditional assessments, students are typically given several choices (e.g., a,b,c or d; true or false; which of these match with those) and asked to select the right answer. In contrast, authentic assessments ask students to demonstrate understanding by performing a more complex task usually representative of more meaningful application.

**Contrived to Real-life:** It is not very often in life outside of school that we are asked to select from four alternatives to indicate our proficiency at something. Tests offer these contrived means of assessment to increase the number of times you can be asked to demonstrate proficiency in a short period of time. More commonly in life, as in authentic assessments, we are asked to demonstrate proficiency by doing something.

**Recall/Recognition of Knowledge to Construction/Application of Knowledge:** Well-designed traditional assessments (i.e., tests and quizzes) can effectively determine whether or not students have acquired a body of knowledge. We *are* often asked to recall or recognize facts and ideas and propositions in life, so tests are somewhat authentic in that sense. However, the demonstration of recall and recognition on tests is typically much less revealing about what we really know and can do than when we are asked to construct a product or performance out of facts, ideas and propositions. Authentic assessments often ask students to analyze, synthesize and apply what they have learned in a substantial manner, and students create new meaning in the process as well.

**Teacher-structured to Student-structured:** When completing a traditional assessment, what a student can and will demonstrate has been carefully structured by the person(s) who developed the test. A student's attention will understandably be focused on and limited to what is on the test. In contrast, authentic assessments allow more student choice and construction in determining what is presented as evidence of proficiency. Even when students cannot choose their own topics or formats, there are usually multiple acceptable routes towards constructing a product or performance.

**Indirect Evidence to Direct Evidence:** Asking a student to write a critique should provide more direct evidence of that skill than asking the student a series of multiple-choice, analytical questions about a passage, although both assessments may be useful.

### **Authentic Assessment (AA) Complements Traditional Assessment (TA)**

But a teacher does not have to choose between AA and TA. It is likely that some mix of the two will best meet your needs. To use a silly example, if I had to choose a chauffeur from between someone who passed the *driving* portion of the driver's license test but failed the *written* portion or someone who failed the driving portion and passed the written portion, I would choose the driver who most directly demonstrated the ability to drive, that is, the one who passed the driving portion of the test. However, I would *prefer* a driver who passed both portions. I would feel more comfortable knowing that my chauffeur had a good knowledge base about driving (which might best be assessed in a traditional manner) and was able to apply that knowledge in a real context (which could be demonstrated through an authentic assessment).

Source: <http://jfmuller.faculty.noctrl.edu/toolbox/whatisit.htm>

## **Interpreting assessment results**

### **Criterion-referenced and standards-based interpretations**

Student's score is compared to clear description of specific tasks a student can perform as described in objectives, learning outcomes which specify the standards of mastery

No need to consider other students' scores

Most useful when test designed for this purpose has:

- set of clearly stated learning objectives

- enough items to infer degree of mastery or non-mastery of that domain

- items selected to actually measure that domain

## **Norm-Referenced Interpretation**

Student's score is compared to those of other students (in a norm group)

Norm group is carefully defined

no need to look at level of mastery

Test scores are interpreted with a norm-referenced or criterion-referenced interpretation, or occasionally both. A norm-referenced interpretation means that the score conveys meaning about the examinee with regards to their standing among other examinees. A criterion-referenced interpretation means that the score conveys information about the examinee with regards a specific subject matter, regardless of other examinees' scores. (Source: [http://en.wikipedia.org/wiki/Test\\_score](http://en.wikipedia.org/wiki/Test_score))

## **Item difficulty or P: the percentage of students who correctly answered an item. Also called Difficulty Index or Facility Index**

Also referred to as the p-value

Ranges from 0% to 100%, or more typically written as a proportion 0.00 to 1.00

The higher the value, the easier the item.

P-values above 0.90 indicate very easy items that you should not use in subsequent tests. If almost all students responded correctly, an item addresses a concept probably not worth testing.

P-values below 0.20 indicate very difficult items. If almost all students responded incorrectly, either an item is flawed or students did not understand the concept. Consider revising confusing language, removing the item from subsequent tests, or targeting the concept for re-instruction.

Facility Index / Difficulty Index is arrived at by dividing the number of examinees who got the item correctly. E.g. - If 10 out of 10 examinees got the test item correctly, that is  $10/10 = 1$ , the difficulty or facility index is 1. If out of 10 students only 2 got the item correctly, the difficulty/facility index is .20. If 6 out of 10 examinees got the item correctly, the difficulty index is .60. THE EASIER THE ITEM THE CLOSER THE

DIFFICULTY INDEX TO 1. The FARTHER THE DIFFICULTY INDEX FROM 1, THE MORE DIFFICULT THE ITEM IS.

### **Discrimination Index**

A good test item is supposed to discriminate between the upper and the lower group.

The upper group and the lower group are arrived at by arranging the corrected test papers from the highest to the lowest score. The upper 27% is the upper group while the 27% scores from below is the lower group.

It is a natural expectation for the upper group to get the item correctly in contrast to the lower group. If more from the upper group got the item correctly than from those in the lower group, the test item will have POSITIVE DISCRIMINATION INDEX. If more from the lower group got the item correctly than those from the upper group contrary to what was expected, the test item has NEGATIVE DISCRIMINATION INDEX. If the same number of examinees from the upper and lower groups got the item correctly the test item has no (ZERO) discrimination index.

If you have a negative Discrimination index, means that students with less skills and knowledge overall, are getting it right more often than those who the test says are better overall.

In other words, the better you are, the more likely you are to get it wrong.

### **What could account for that?**

Two possibilities:

- Usually means an ambiguous question
  - that is confusing good students, but weak students too weak to see the problem
  - What should a teacher do? Look at question again, look at alternatives good students are going for, to see if you've missed something.

OR:

- or it might be off topic
- --> something weaker students are better at (like rote memorization) than good students

**Discrimination Index = % Correct in High Group - % Correct in Low Group**

No. of students who got the item correctly (-) no. of students in the lower group who got the item correctly divided by the number of students in the upper group.

**Item analysis** - When you do item analysis you analyze the quality of test items and so you compute difficulty/facility index, discrimination index, attractiveness of the options if it were multiple choice.

## **Score Distribution**

**Normal curve – Bell curve – Mesokurtic – Symmetrical**

With drawing

In a **normal curve distribution**, the mean is equal to the median and to the mode.

Sometimes the score distribution does not follow the bell curve. The score distribution gets **skewed or asymmetrical**. **Skewed to the left** (negatively skewed) or **skewed to the right** (positively skewed). The skewness is shown by the tail. If the tail of the distribution is on the left and the mountain of scores is on the right then you have a negatively skewed distribution.

**Negatively skewed** distribution suggests high scores, good performance, brilliant students, good teaching.

**Positively skewed** score distribution implies low scores, non-performing students, ineffective teaching.

**Measures of central tendency** – scores tend to congregate; how scores clustered

**Mean** – average

**Mode** – the most frequent score

**Median** – midpoint of a score distribution

When scores are extremely high and low, the median is a more appropriate measure for central tendency.

**Measures of Variability** - how spread the scores are

**Range** – the difference between the highest and the lowest score

**Inter-quartile range** - The inter-quartile range is a measure that indicates the extent to which the central 50% of values within the data set are dispersed. It is based upon and related to, the median.

**Standard deviation** - **Indicates how tightly the values in the dataset are bunched or clustered around the mean value. It is the most robust and widely used measure of dispersion since, unlike the range and inter-quartile range, it takes into account every variable in the data set.**

**Sources:**

<http://www.sfsu.edu/~testing/MCTEST/guidelines.html>

<http://www.studygs.net/teaching/tsttak3a.htm>

<http://www.utexas.edu/academic/ctl/assessment/iar/students/plan/method/exams-mchoice-write.php>

<http://www.studygs.net/teaching/tsttak2a.htm>

[http://oct.sfsu.edu/assessment/measuring/htmls/essay\\_tests.html#dos](http://oct.sfsu.edu/assessment/measuring/htmls/essay_tests.html#dos)

[http://som.unm.edu/ume/ted/pdf/ed\\_dev/assess\\_dev\\_essay.pdf](http://som.unm.edu/ume/ted/pdf/ed_dev/assess_dev_essay.pdf)

<http://ed400fall09.wikispaces.com/Essay+-+Extended+Response>

## **Assigning Students Marks/ Grading**

**DepEd Order 73, s. 2012**

**Assessment shall be used primarily as a quality assurance tool :**

to track student's progress in the attainment of standards

promote self-reflection and personal accountability for one's learning

and provide a basis for the profiling of student performance.

To determine students' attainment of standard in terms of contents and performance.

To track students' progress in learning.

To gather data for appropriate intervention. Traditional and Non-traditional Assessment Compared

#### TRADITIONAL

- Testing
- Paper-pencil
- Multiple-choice type
- Single correct answer
- Summative
- Outcomes only
- Skill-focused
- Isolated facts
- Decontextualized task
- External evaluation

#### NON-TRADITIONAL

- Alternative assessment
- Performance type
- Supply
- Many correct answers
- Formative
- Process and outcomes
- Task-based
- Application of knowledge
- Contextualized tasks
- Student self evaluation

Assessment **FOR** Learning – **FOR**mative Assessment

Assessment **AS** Learning – Self-assessment

Assessment **OF** Learning – Summative assessment

## Principles of Assessment

### Holistic

- Diagnostic (assessment for learning)
- Formative/Developmental (assessment for and assessment as learning)
- Summative (assessment of learning)

### Standards-based and outcomes-based

- ensures that teachers will teach to the standards and students will meet or even exceed standards. Evidence of learning is the students' attainment of content and performance standards. The learning outcomes are defined in the standards.

### **Learning outcomes by level**

The learning outcomes come in different levels - knowledge, process or skill, understanding and products and performances.

### **Levels of Learning Outcomes**

Knowledge – the substantive content of the curriculum, the facts and information that the student acquires

Process – skills or cognitive operations that the student performs on facts and information for the purpose of constructing meanings or understandings

Understandings – enduring big ideas, principles and generalizations inherent to the discipline, which may be assessed using the facets of understanding or other indicators of understanding which may be specific to the discipline

Products/Performances – real life application of understanding as evidenced by the student's performance of authentic tasks

### **4. Multiple measures**

Learning in different levels is likewise assessed thus assessment includes multiple measures. The assessment of student performance makes use of multiple measures and tools such as quizzes, participation, projects, periodical tests and homework to assess learning in different levels from acquisition of knowledge to actual performance which usually leads to a product.

**Knowledge** – This level may be assessed by the use of traditional measures (e.g. paper-and-pencil tests using multiple choice, true-false or matching type of tests if the intention is to find out students' knowledge of specific facts and information.

**Process and Skills** – Assessment may focus on how logically, analytically, or critically students make sense of or process the information or use it. This may be assessed by asking learners to outline, organize, analyze, interpret, translate, convert or express the information in another form or format, draw analogies; construct graphs, models, flowcharts and mind maps, or graphic organizers; or transform a textual information into a diagram; draw or paint pictures or do role plays to express creatively their sense of the facts and information.

**Understanding** – This level focuses on the meanings that students themselves make or develop. The assessment should be able to draw from the students the meaning/s they have made or their own understanding. This can be expressed in any of the six facets of understanding- explain, interpret, apply, give perspective, show empathy and self-knowledge.

When students understand they can:

1. explain concepts, principles, and processes by using their own words, teaching them to others, justifying their answers and showing their reasoning
2. interpret by making sense of data, text, and experience through images, analogies, stories and models
3. apply by effectively using and adapting what they know in new and complex contexts
4. demonstrate perspective by seeing the big picture and recognizing different points of view
5. display empathy by perceiving sensitively and putting one's self in someone else's shoes
6. have self-knowledge by showing meta-cognitive awareness, using productive habits of mind and reflecting on the meaning of the learning and experience

## Grading System

1. K to 12 makes use of the descriptive rating system. This does not mean that teachers will no longer compute. The descriptive rating will be based on a numerical value which is arrived at after summing up the results of the student's performance on the various levels of assessment

Level of Proficiency	Equivalent Numerical Value
Beginning	74% and below
Developing	75 – 79%
Approaching Proficiency	80 – 84%
Proficient	85 – 89%
Advanced	90 and above

## Meaning of Descriptive Grades

**B** for Beginning – This means that the student “struggles with his/her understanding; pre-requisite and fundamental knowledge and/or skills have not been required or developed adequately to aid understanding

**D** for Developing – Students given with “D” are those that “possess the minimum knowledge and skills and core understandings but needs the help throughout the performance of authentic tasks”

**AP** for Approaching Proficiency – Students that have “developed the fundamental knowledge and skills and core understandings and with little guidance from teacher and/or with some assistance from peers;; can transfer these understandings through authentic performance tasks.”

**P** for Proficient – Students that have “developed fundamental knowledge and skills and core understandings and can transfer them independently through authentic performance tasks.”

**A** for Advanced – Students “exceed the core requirements in terms of knowledge, skills, and understandings and can transfer them automatically through authentic performance tasks.”

(Source : DepEd Order 73,s.2012 )

#### **REFERENCES:**

- Baker E. L., “The role of domain specifications in improving the technical quality of performance assessment” (CSE Tech. Rep.). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing, 1992.
- Barton P. E., and Coley R.J., “Testing in America’s Schools Policy Informational Report”, ETS, 1995, 46 p.
- Bhatnagar, G and Sharma A. “Learn Computing: A Game Plan”, Proceedings of the F13 conference, August 1999, NIIT Ltd., New Delhi.
- Bryant D., “A Comparison of Multiple Choice versus Alternative Assessment: Strengths and Limitations”, The New York State Education Department, Albany, New York, 1996 (available at <http://www.nysed.gov/rscs/rschmult.html>)
- Cartwright, Rebecca et al. 2009. Student Learning Outcomes Assessment Handbook. Maryland: Montgomery College.
- Darling – Hammond, Linda and John Bransford (eds.) 2005. Preparing Teachers for a Changing World. Ca: John Willey and Sons.
- DepEd Order 73,s.2012
- Dunbar S. B., Koretz D., & Hoover H. D. (1991). “Quality control in the development and use of performance assessments”, *Applied Measurement in Education*, 4, 1991, 289-304.
- Dunbar, S. B. et al. 1991. “Quality Control in the Development and Use of Performance Assessments,” *Applied Measurement in Education*. 4. 289-304.
- Guba, Egon and Yvonne Lincoln. 2011. *Fourth Generation Evaluation*, Ca: Saga Publishing Co.
- Hart, Diane. 1994. *Authentic Assessment: A Handbook for Educators*. Ca: Addison – Willey Publishing Co.
- Herman J., “Research in cognition and learning: Implications for achievement testing,” In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 154-165) Englewood Cliffs, NJ: Prentice Hall, 1991.
- Hernon, Peter and Robert Dugan (eds.). 2004. *Outcomes Assessment in Higher Education*. Westport: Libraries Unlimited.
- Johnson, K.E. 1996. “Portfolio Assessment in Second Language Teacher Education,” *TESOL Journal*, Vol. 2
- Lim, R.L. et al. “Complex Performance – Based Assessment: Expectations and Validation Criteria,” *Educational Researcher*, 20(8), 15-21.
- Linn R. L., Baker E. L., & Dunbar S. B., “Complex Performance-based assessment: Expectations and validation criteria”, *Educational Researcher*, 20(8), 1991, 15-21.

- Lipman M., "Some Thoughts on the Formation of Reflective Education", In *Teaching-Thinking Skills: Theory and Practice*, pp. 151-161. Edited by J.B. Baron and R. J. Sternberg. New York: W. H. Freeman, 1987.
- Lukhele R., Thissen D. and Wainer H, "On the Relative Value of Multiple Choice Constructed Response and Examinee Selected Items on Two Achievement Tests", *Journal of Educational Measurement*. 31 (3), 1994, 234-250
- Navarro, Rosita L. and Rosita G. Santos, (2013) "Authentic Assessment of Student Learning Outcomes" 2nd Edition. Quezon City, Philippines: Lorimar Publishing, Inc.
- Mehrens W. A., "Using performance assessment for accountability purposes" *Educational Measurement: Issues and Practice*, 11(1), 1992, 3-20.
- Mueller, Jon. 2011. *Authentic Assessment*, Nashville, Ill.: Northcentral Colleges.
- O'Malley, J.M. and L. Valdez Pierce, 1996. *Authentic Assessment for English Language Learners*. Ca.: Addison-Wesley Publishing Co.
- Santos, Rosita G. (2012) "Advanced Methods in Educational Assessment and Evaluation", 2nd Edition. Quezon City, Philippines: Lorimar Publishing, Inc.
- Schoenfeld A. H., "Mathematical Problem Solving", Academic Press, Orlando FL, 1985.
- Seldon A., Seldon J., "What does it take to be an expert problem solver", *Research Sampler, MAA Online*, No. 4. (1997); (available at [www.maa.org/t\\_and\\_l/sampler/research\\_sampler.html](http://www.maa.org/t_and_l/sampler/research_sampler.html)).
- Shavelson R. J., Baxter G. P., & Pine J., "Performance assessments: Political rhetoric and measurement reality", *Educational Researcher*, 21(4), 1992, 22-27.
- Shulman, L.S. and J. Shulman. 2004. "How and What Teachers Learn: A Shifting Perspective," *Journal of Curriculum Studies*. 36 (2) pp. 257-271.
- Starch D., & Elliot E. C., "Reliability of grading high school work in English" *School Review*, 20, 1912, 442-457.
- Starch D., & Elliot E. C., "Reliability of grading high school work in mathematics", *School Review*, 21, 1913, 254-259.
- Tuckman B. "The Essay Test: A Look at the Advantages and Disadvantages" *NASSP Bulletin*, 77, 1993, pp. 20-26.
- Wiggins, Grant and McTighe. 1998. *Understanding by Design*. New Jersey: Prentice Hall.
- Wolf, K. and Y. Siu-Runyan. 1996. "Portfolio Purposes and Possibilities," *Journal of Adolescent and Adult Literacy*. 40: 1, pp. 30-36.
- Zaremba S., and Schultz M., "An Analysis of Traditional Classroom Assessment Techniques and Discussion", ED 365404, 1993, p.13.
- Zaremba, S. and M. Schultz. 1993."An Analysis of Traditional Classroom Assessment Techniques and Discussion," ED 365404. p. 13.
- <http://caacentre.lboro.ac.uk/dldocs/Bluepaper1.pdf>
- <http://www.brookes.ac.uk/services/ocsltd/resources/assessment/principles.html>
- <http://www.qub.ac.uk/directorates/AcademicStudentAffairs/CentreforEducationalDevelopment/AssessmentFeedback/Assessment/Principlesofassessment/>
- <http://arc.boardofstudies.nsw.edu.au/go/k-6/principles-of-assessment-for-learning/>
- <http://www.cshe.unimelb.edu.au/assessinglearning/05/>
- <http://www.wvcc.wy.edu/assessment/assessmentresults.htm>
- <http://citl.indiana.edu/resources/teaching-resources1/teaching-handbook-items/test-construction.php>
- <http://www.cte.cornell.edu/documents/Test%20Construction%20Manual.pdf>
- <http://suen.educ.psu.edu/~hsuen/pubs/Taiwan.pdf>
- <http://kidedotals.com/book/export/html/493>

[http://teacher.scholastic.com/products/sri\\_reading\\_assessment/pdfs/Interpreting\\_Assessment\\_Results.pdf](http://teacher.scholastic.com/products/sri_reading_assessment/pdfs/Interpreting_Assessment_Results.pdf)

<https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php>

[http://www.icoachmath.com/math\\_dictionary/Measures\\_of\\_Central\\_Tendency.html](http://www.icoachmath.com/math_dictionary/Measures_of_Central_Tendency.html)

<http://www.sfsu.edu/~testing/MCTEST/guidelines.html>

<http://www.studygs.net/teaching/tsttak3a.htm>

<http://www.utexas.edu/academic/ctl/assessment/iar/students/plan/method/exams-mchoice-write.php>

<http://www.studygs.net/teaching/tsttak2a.htm>

[http://oct.sfsu.edu/assessment/measuring/htmls/essay\\_tests.html#dos](http://oct.sfsu.edu/assessment/measuring/htmls/essay_tests.html#dos)

[http://som.unm.edu/ume/ted/pdf/ed\\_dev/assess\\_dev\\_essay.pdf](http://som.unm.edu/ume/ted/pdf/ed_dev/assess_dev_essay.pdf)

<http://ed400fall09.wikispaces.com/Essay+-+Extended+Response>

**ANSWER KEY (ASSESSMENT)**

1 A

2 A

3 B

4 C

5 A

6 A

7 B

8 C

9 A

10 A

11 C

12 B

13 A

14 D

15 A

16 A

17 D

18 A

19 B

20 C

21 A

22 A

23 B

24 A

25 C

26 C

27 A

28 C  
29 A  
30 A  
31 A  
32 A  
33 C  
34 D  
35 C  
36 A  
37 A  
38 C  
39 A  
40 A  
41 C  
42 D  
43 D  
44 D  
45 B  
46 D  
47 D  
48 A  
49 A  
50 D  
51 C  
52 B  
53 D  
54 B  
55 A  
56 A  
57 A  
58 A  
59 A  
60 B  
61 D  
62 D  
63 A  
64 C  
65 D  
66 A  
67 C  
68 D  
69 D

70 A  
71 C  
72 A  
73 B  
74 A  
75 A  
76 A  
77 A  
78 C  
79 A  
80 A  
81 C  
82 A  
83 A  
84 D  
85 A  
86 A  
87 D  
88 A  
89 D  
90 A  
91 A  
92 D  
93 A  
94 C  
95 D  
96 D  
97 B  
98 A  
99 A  
100 A  
101 A  
102 A  
103 D  
104 A  
105 A  
106 C  
107 D  
108 A  
109 B  
110 D  
111 C

112 B  
113 D  
114 D  
115 A  
116 D  
117 A  
118 A  
119 D  
120 A  
121 C  
122 A  
123 A  
124 B  
125 D  
126 C  
127 C  
128 D  
129 D  
130 D  
131 D  
132 B  
133 A  
134 D  
135 A  
136 A  
137 A  
138 A  
139 C  
140 A  
141 A  
142 A  
143 C  
144 C  
145 C  
146 A  
147 A  
148 A  
149 C  
150 D  
151 D  
152 A  
153 C

154 C  
155 B  
156 B  
157 A  
158 A  
159 C  
160 D  
161 A  
162 A  
163 A  
164 A  
165 D  
166 C  
167 A  
168 B  
169 C  
170 A  
171 A  
172 D  
173 A  
174 A  
175 A  
176 A  
177 A  
178 A  
179 C  
180 D  
181 B  
182 A  
183 A  
184 C  
185 A  
186 A  
187 A  
188 A  
189 C  
190 A  
191 A  
192 D  
193 A  
194 A  
195 A

196 A  
197 A  
198 A  
199 D  
200 A  
201 C  
202 C  
203 D  
204 A  
205 B  
206 C  
207 D  
208 A  
209 A  
210 A  
211 C  
212 B  
213 A  
214 A  
215 A  
216 D  
217 A  
218 A  
219 D  
220 D  
221 C  
222 B  
223 A  
224 B  
225 A  
226 C  
227 B  
228 A  
229 A  
230 A  
231 A  
232 D  
233 A  
234 C  
235 D  
236 C  
237 C

238 A

239 B

240 A