# DESIGNING AND DEVELOPING SEMESTER TEST AT SMP N 1 SINTUK TOBOH GADANG PADANG PARIAMAN

## Susi Karmila

Universitas Negeri Padang e-mail: susikarmila292@gmail.com

Abstract: The purpose of this research was to determine how an English teacher design and evaluate the semester test. This research was done in SMP N 1 Sintuk Toboh Gadang. One of the ways through which feedback can be obtained from the learners on what the teachers had taught them is an evaluation. Students' achievement in a particular course of study can be determined through evaluation. This work observed various aspects of ordinary level English in which students' achievement are often assessed with a view to assisting students in overcoming problems often encountered during such tests and evaluations. Based on the results of observation and interviews with English teachers at the school, explained that for designing the semester test English language teachers should be used to assess the degree to which students have learned what they should have learned and provided a vehicle for providing students feedback about how well they have learned the material. To solve the problem English teacher design and evaluate the semester test by using the grating to help students.

Keywords: Designing and Developing, Semester Test.

### 1. INTRODUCTION

planning In the and development of an English language proficiency assessment be administered to international testtakers, the same general principles of good assessment practices used with other types of assessments apply. Most importantly, the purposes of an assessment must be clearly specified. It is also essential to develop a precise and explicit definition of the construct the assessment is intended to measure. The underlying theoretical rationale for the existence of the construct should be articulated. An assessment that is built on a strong theoretical foundation is one that is more likely to lead to valid interpretations of test scores.

In addition, a clear definition of the construct being measured can help clarify the skills associated with that construct. This enables test developers to create tasks for an assessment that will best engage the test-takers skills and reflect the construct of interest. A good standardized test is the product of a through the process of empirical research and development that may extend beyond simply the establishment ofthe standard. Evaluation is one of the ways through which feedback can be obtained from the learners on what their teachers had taught them.

#### **Review of Related Literature**

In education. the term assessment refers to the wide variety of methods or tools that educators use to evaluate, measure, and document academic readiness, learning progress, skill acquisition, educational needs of students. Most elementary and secondary schools around the world have standardized achievement tests to measure students' mastery of the students competencies that have been prescribed grade levels. exit requirements, and entrance to further levels. While assessments are often equated with traditional testsespecially the standardized tests developed by testing companies and administered to large populations of students. Carrol (1968) states that test a psychological or educational test is a procedure designed to elicit certain behavior from which one can make inferences about certain characteristics of an individual.

Evaluation has earlier been described as one of the ways through which feedback can be obtained from the learners what they have been taught. Students' evaluation should be based on the objectives of teaching earlier set before the commencement of teaching. Evaluation according to Adewuyi and Oluokun (2001) is the

process of gathering and interpreting evidence regarding the problems and progress of learners in achieving desirable educational goals. Ogunwuyi (2004), quoting Okpala, Onocha and Oyedeji (1993) presented evaluation as gathering process of information on the attainment of educational objectives, analyzing and fashioning information to aid judgment on the effectiveness of teaching or an educational programme. Therefore, evaluation is necessary to be able to determine the extent to which learning has taken place. It also assists in determining the level of students' understanding and it gives opportunity of rating them accordingly.

Desheng and Vargbese (2013) defined evaluation as the comparison of actual (project) impacts against the agreed strategic plans. According to them, evaluation looks at the, original objectives, at what was accomplished, and how it was accomplished. Students' achievement can be determined through evaluation. In most cases, evaluation of students' achievement is based on the behavioral objectives of instruction or lesson in question. Instructional or behavioral objectives, therefore, provide us with a set of vardsticks for evaluating learning.

There are certain qualities expected of a good language test. These are the characteristics of a good language test. They include, among others, validity, reliability, objectivity, and economy.

- 1. Validity: A good language test should measure what it is supposed to measure. There are different types of validity. These are: (i) face validity (ii) content validity (iii) predictive validity (iv) concurrent validity; and (v) construct validity.
- 2. Reliability: Reliability is the quality of being reliable. Language test reliability is the consistency of a test in measuring what it is supposed to measure. A good language test is expected to be reliable. There are two kinds of test reliability. These are intra-rater and inter-rater reliability. Intra-rater reliability indicates that a scorer's mood, physical environment and psychological state of mind may affect the mark given a particular essay at different times. Inter-rater reliability posits that there tends to be variation in the mark given to a particular essay by two or more scorers (markers).
- 3. Objectivity: This quality of a language test ensures that a test should have one and only one correct answer. Examples of these include the "multiple choice" and "True and False" tests.
- 4. Economy: This quality of a test ensures that the cost of administering a test, the time involved in setting and marking it should be commensurate with the expected result obtained from it. A test that takes much time, energy and costs much to construct cannot be said to be economical.

Six Steps of Developing Standardized Test

1. Determine the purpose and objective of the test

Most standardized tests are expected to provide high practically in administration and scoring without unduly compromising validity. The initial outlay of time and money for such a test significant, but the test is usually designed for repeated use. Is, therefore, we important for its purpose and objective to be stated specifically.

2. Design test specifications

Decisions need to be made on how to go about structuring the specifications of the test. This stage of laying the foundation stones can occupy weeks, month, or even years of effort. To illustrate the designer of test specs, focus on the TOEFL. Because, TOEFL is a proficiency test, the first step in the development process because is to define the construct of language proficiency.

3. Design, select and arrange test tasks/items

The specs act much like a blueprint for determining the number and types of items to be created. The designer would then need to select the passage, the cloze items, the 15 possible fill in words and the 5 inference items. The reading passage would have to adhere to certainly predetermined reading difficulty specs, so one would be forced to work within those parameters.

4. Make appropriate evaluations of different kind of items.

Especially if the classroom-based test is a onetime test. Yet for a standardized multiple choice test that is designed for a commercial market, and administrated a number of times or administrated in a different form, these indices are a must. For other types of response formats namely, only and written responses, different forms of evaluation become important. The principles of practicality and reliability are prominent, along with the concept of the facility.

5. Specify scoring procedure and reporting formats

A systematic assembly of test items in preselected arrangement and sequences, all of which are validated to conform to an expected difficulty level, should yield a test that can then be scored accurately and reported back to test takers and institution efficiently.

6. Perform ongoing construct validation studies

It should be clear that no standardized instrument is expected to be used repeatedly without a rigorous program of ongoing construct validation. All of the tests have examined here include programs of construct validation, especially in view of their need to periodically produce the new form of the test.

One of the key aspects of standards-based assessment is post-assessment feedback. The feedback a student receives from this type of assessment does not emphasize a score, percentage, or statistical average, but information about the

expectations of performance compared to the standard. The purpose of standards-based assessment is to connect evidence of learning to learn outcomes (the standards). standards are explicit and clear, the learner becomes aware of his/her achievement with reference to the standards, and the teacher may use assessment data to give meaningful feedback to students about progress. One of the chief reasons that students' standardized test scores continue to be the most important factor in evaluating a school is deceptively simple. Most educators do really understand why standardized test provides a misleading estimate of school staff's effectiveness.

Evaluation can be defined as the systematic gathering of information for the purpose of making decisions (Weiss 1972). The probability of making the correct decision in any given situation is a function not only the ability of the decision maker but also of the quality of the information upon which the decision is based. Everything else is equal, the more reliable and relevant the information.

A primary concern in education is whether students attain the general. Indeed, a common way of thinking about evaluation is in term of student achievement at the end of a course of instruction in comparison to the beginning of course objectives. Evaluation need not wait until instruction has ended. In fact, it is

advisable to begin evaluation before instruction actually starts. Waiting until an instruction is finished minimizes the beneficial effects of evaluation on instruction and student learning. Once this has been done, the next step is the collection of information pertinent to this decision. This aspect of evaluation is referred to an assessment.

Assessment information is seldom useful by itself. It must be interpreted put into context before it is meaningful. This accomplished by comparing it with some desired state of affairs, goals, or other information that you have that is relevant to your decision. Once this is done, an interpretation of the information is possible, and finally, a decision can be made about how to piece. This fourstep process identifying purposes, collecting information, interpreting it, and making the decision characteristic of every choice as well as professional decisions teachers make when they are planning instructions and engaged in teaching. Five Steps in Designing an Effective Test as Classroom Language Assessment

- Determining the purpose of a test First, new and innovative testing formats take a lot of effort to design and a long time to refine through trial and error. Second, traditional testing techniques can with a little creativity conform to the spirit of an interview, communicative language curriculum. The first and perhaps most important step in designing any sort of classroom assessment is to step back and consider the overall purpose of the exercise that your students are about to perform.

- Designing clear, unambiguous objectives

In addition to knowing the purpose of the test you're creating, you need to know as specifically as possible what it is you want to test. This no way to approach a test. Indeed, begin taking a careful look at everything that you think your students should know or be able to do base on the material that the students are responsible for. In other words, examines the objectives of the unit you are testing. Remember that curriculum should have every appropriately framed assessable objectives that are objectives that are stated in term of overt performance by students.

- Drawing up test specifications
  Test specifications for classroom use
  can be an outline of your test what it
  will look like. Think of your test specs
  as a blueprint for the test that includes
  the following:
- a. A description of its content
- b. Item types
- c. Tasks
- d. Skill to be included
- e. How the test will be scored
- f. How it will be reported to students. Grammar unit test that design might comparison the following four sequential steps:
  - a. A broad outline how the test will be organized

- b. Which of the eight subskills you will test
- c. What the various tasks and item types will be
- d. How result will be scored, reported to students, and used in a future class

Midterm essay by the specification assessment the following:

- a) Provide clear direction
- b) Write a prompt on either a narrative or a description essay
- c) The prompt must be on a familiar topic that students will
- d) Assign an expectation of a full page, handwritten, and no more than two full pages
- e) Students will be allowed a 90 minutes time limit
- f) In the prompt, include evaluation criteria
- g) The final grade is to include four subscores content, organization, rhetorical discourse and grammar

## 1] Devising test items

At this point, it is important to note that test development is not always a clear, linear process. In reality, test design usually involves a number of loops as discover problems and another shortcoming. There is kind of item in devising test items include midterm essay, listening or speaking final exam, listening comprehension section and oral presentation.

2] Designing multiple-choice items

First, we need to turn out attention to some important principles and tips for designing multiple-choice tests.

Multiple choice items, which may on the surface appear to be simple items to construct, are actually very difficult to design correctly. The two principles that stand out in support of multiple choice formats are, of course, practicality and reliability. Multiple choice items are all receptive or selective response items in that the test taker chooses from set responses rather than creating a response. perspective item types include true or false questions and matching lists. Consider the following four guidelines for designing multiple-choice items for both classroom-based and large-scale situation:

- a. Design each item to measure a single objective
- b. State both an option as simply and directly as possible
- c. Make certain that the intended answer is clearly the only correct one
- d. Use item indices to accept, discard, or revise items

#### 2. METHOD

This research conducted in SMP N 1 Sintuk Toboh Gadang Padang Pariaman. Collecting data through interview and observation with an English teacher at the school. Data were analyzed through descriptive.

### 3. FINDINGS

SMP N 1 Sintuk Toboh Gadang, designing effective test question by understanding the problem that classroom face. English teacher at

SMP N 1 Sintuk Toboh Gadang designing ad evaluate the classroom test. Classroom test serve two primary purposes, they are should be used to assess the degree to which students have learned why they should have learned and provided a vehicle for providing students feedback about how well they have learned the material. If the test not well designed, then neither of these goals is served well because the test results do not reflect students actually know.

Based on interviews with the English teacher at the school explained that, when students perform poorly on tests, we often assume that poor performance is due to lack of student ability or effort studying for the test. The tendency is to "blame the student" and then lower both our perceptions. So, the solution of this problem according to English teacher at this school is when students are not performing well on tests, the first factor to consider is not the student's capabilities or efforts studying, but rather the quality of the test itself. Often minor improvements assessment procedures can mean the difference between success and failure for many students. Aspects of a test that can be clarified for the benefit of all students include writing the goal is unambiguous provide clear, to directions and test questions that are free of convoluted, confusing language and unfamiliar vocabulary.

To overcome the problem, to make the exam by English teacher

based on the grating matter. Grating (test blueprint or table of the specification) a description of the competence and the material to be tested. The purpose of the grating is to determine the scope of and as a guide in writing about. Good grating must meet the following requirements,

- a. The grating should be able to represent the content of the syllabus/curriculum or materials that have been taught appropriately and proportionately.
- b. The components described in a clear and easily understood.
- c. The material can be made because of any inquiries.

Indicators in the grating are a guideline in formulating a matter of choice. The formulation of indicators about a part of the preparation of the lattice. To formulate indicators right, teachers must pay attention to the material to be tested, the indicators of learning, basic competence, and competence standards. A good indicator is formulated briefly and clearly.( grating on the attachment).

According to English teacher in order for a matter which is prepared by each teacher generating material quiz / test is valid and reliable, then it should do the following steps, namely: (1) determine the purpose of the test, (2) determining the competencies will be tested, (3) determine the material being tested , (4) specify deployment items based on competence, materials and shapes assessment (written test: multiple choice, analysis, and testing

practices), (5) make up the lattice, (6) write items, (7) to validate items or examine qualitatively, (8) to assemble the item into the test device, (9) develop guidelines for score (10) test items, (11) the analysis of items quantitatively from the empirical data test results, and (12) repair matter based on the analysis.

Test directions are a critical and often overlooked aspect of test construction. The rest of the test is unimportant if the student does not clearly understand the directions. A talented item writer can construct multiple-choice items that require not only recall of knowledge but also comprehension, interpretation, applications, analysis, or synthesis to arrive at the correct answer. Keep as much of the item as possible in the stem (the question) and keep the answer and distracters as short as possible. This makes for less reading and less chance for confusion. Include in the stem only the material needed to make the problem clear and specific. The test should assess the student's knowledge of the material, not the ability of a student to locate a question from a wordy passage. These questions often require less knowledge for testwise students than traditionally worded questions and are confusing to students with less ability. Arrange the answer and distracters vertically on the page.

### 4. CONCLUSION

Evaluation is a process that can guide educational decision making and that there are four aspects to this process (identifying the purpose for collecting evaluation, information, interpreting information and making the decision). We describe a strategy for making decisions that consist of comparing instructional objectives, plans, practices, input factors and learning outcome with one another. The abilities we want to measure, the test methods that we use have an important effect on test performances. If we consider the variety of testing technologies that are used in language test and the ways in which these techniques vary, it is obvious that test methods vary along a large number of dimensions. If we understand the ways in which test methods influence test performance, therefore it is necessary to examine the various dimension or facet of test methods.

Assessment requires students to perform tasks that were included in the previous lesson and represent the objectives of the unit on which the assessment is based. Students will judge a test to be valid if: Directions are clear, the structure of the test is organized logically, its difficulty level is appropriately pitched, the test has no "surprises" and Timing is appropriate. Biased for best: a term that goes a little beyond how the student views the test to a degree of strategic involvement on the part of student and teacher in preparing for, setting up, and following up on the test. The design of an effective test should point the way to beneficial washback. A test that achieves content validity demonstrates relevance to the curriculum in question and sets the stage for washback

## **REFERENCES**

- Adewuyi, J.O. and Oluokun, O. (2001). Introduction to test & measurement in education.

  Oyo,Odumatt Press and Publishers.
- Bachman, Lyle F. 1990. Fundamental Consideration in Language Testing. Oxford: Oxford University Press.

- Brown, H. Douglas. 2004. Language
  Assessment: Principles and
  Classroom Practices. New
  York: Longman.
- Brown, H. Douglas and Priyanvada Abeywickrama. 2010. Language Assessment: Principles And Classroom Practices. NY: Person Education.
- Desheng, C. and Varghese, A. (2013). Testing and evaluation of language skills. In *IOSR Journal of Research*.
- Genesee, Fred and Upshur, John A.
  2002. Classroom-Based
  Evaluation in Second Language
  Education. Cambridge:
  Cambridge University Press.
- McNamara, Tim. 2000. Language Testing. Oxford: Oxford University Press.