

Definition of test types of test education essay



This chapter provides an overview of the theories related to the study concerning with items test analysis. To be more specific, the review of related literature discusses about: definition of test, types of test, achievement test, summative test, English test of Tes Kendali Mutu, aspects in developing a test, validity in language testing, content validity, the school based curriculum, items test analysis, difficulty level, discrimination power, distractor function, and theoretical framework.

Definition of Test

A test supposed to be able to measure learning outcome which distinguish the every single student's ability between students already mastered and not yet the learning material. Therefore, testing is one of the powerful tools to measure students' abilities as well as enhance their attitudes towards learning. This notion is supported by Hughes (2003) stated that a test is a tool to measure language proficiency of students. Brown (2004: 3) stated that a test is a method of measuring a person's ability knowledge, or performance in a given domain. In the same line, Anthony J Nitko (1983: 6) defined test is systematic procedure for observing and describing one or more characteristics of person with the aid of either a numerical or category system. Based on Cronbach in Azwar (2005) defined a test is a systematic procedure for observing a person's behavior and describing it with the aid of a numerical scale or category system.

In short, a test as an instrument of evaluation is a systematic procedure of description, collection and interpretation in order to measure the test taker's achievement ability, knowledge, and performance what they have been learned in learning process and to get a value judgment. The purpose of a

test is able to give the valid information on the students' abilities and knowledge. Hence, the successfulness of the teaching and learning can be seen in the test's results.

Types of Test

A test can be classified based on the types of information they provide.

Based on Wilmar Tinambunan in his book with the entitled Evaluation of Students Achievement (1998: 7-9), the four types of test are placement test, formative test, diagnostic test, and summative test. The language tests have different kinds of purposes. Baily (1998) hold that there are eight kinds of language assessment. They are aptitude test, language dominance test, proficiency test, admission test, placement test, diagnostic test, progress test, and achievement test.

Based on different classification of several types of tests described from any experts of language testing depending on variable purposes in testing if related to this study that the types of tests that will be reviewed and used is achievement test and summative test.

Achievement Test

Achievement test are directly related to language courses, their purposes being to establish how successful individual students, groups of students, or the courses themselves have been achieving objectives. In the book of Dictionary of Language Teaching and Applied Linguistics that written by Jack C. Richard (1992: 3), he said that an achievement test is a test which measures how much of a language someone has learned with reference to particular course of study or program of instruction. In the same line, Brown

(2004) defined an achievement test is to see how far students achieve materials addressed in a curriculum within a particular time frame. It means that an achievement test is to measuring students' achievement learning outcome which is administered at the end of course of study. The scope of test content must that represents the course they are concerned.

Summative Test

Summative test is an achievement test administrated at the end of a course or unit of instruction. Brown (2004) mentioned that final exam in a course is example of summative test. Summative test is intended to show the standard that the students have now reached in relation to other students at the same stage. Thus summative test is conducted to find out overall learning outcome after set of learning units program already done.

Summative test is commonly designed based on subject matter that had been already taught for one semester. Consequences of a test that emphasize the overall learning outcomes is its substance of scope involve all of material have been delivered. The result of summative test is to determine successful in learning and teaching, whether or not the students are able to follow the higher level of next instruction program, and students' progress. In the final, the level of success rate is stated with a score or value which written in the form of report card.

English test of Tes Kendali Mutu

Tes Kendali Mutu is a name of appliance evaluates administered at SMKN 26 Jakarta after teaching and learning process for measuring how far purpose of students' learning and teachers' teaching achievement reached. As in final semester examination, according to Regulation Minister of National

<https://assignbuster.com/definition-of-test-types-of-test-education-essay/>

Education No. 20/2007, it is carried out to measure students' competency achievement at the end of the semester. It can be said that Tes Kendali Mutu is administrated at each final in learning and teaching process in the period one semester and its scope includes all the indicators that represent all the basic competence at that semester. All groups of subject are tested in order to evaluate learning outcome including an English subject. The format of English test of Tes Kendali Mutu of first grade school year 2011/2012 at SMKN 26 Jakarta is written test with objective multiple choice items. Total number English items test of Tes Kendali Mutu is 50 items test which including listening and reading skill. Listening skill consists of 15 items test number and reading skill consists of 35 items test number. In reading skill, there are three reading section which consists of first section was 15 incomplete short dialogue items; second section was 5 items error recognition; and third section was 15 questions of reading comprehension which divided into different 5 of reading texts.

Aspects in Developing a Test

In order to know criteria of a good test for measuring students' ability has been reached in learning process, a test as an instrument of evaluation has to meet requirements the validity, reliability, and practicality. As stated by Brown (2004), there are three important aspects should reflect in a test, namely validity, reliability, and practicality. Gronlund (1998) pointed out that validity is the extent to which inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of the assessment. Validity has to do with the information that the uses in class, it has to be appropriate for the student's level, the purpose of the class, and if

the meaning of the materials used in class are for the students. In term of achievement test, it can be valid if it is accurate, authentic or valid has been able to measure students' learning outcomes achieved, after they achieved the learning process in a certain period. Reliability is consistent and dependable. A test is called reliable if the same test is given to the same students on two different occasions and its tests should yield similar results. Practicality means an effective test. A test should be efficient and easy to be implemented. Since the focus of analysis was merely on one aspect of language test, it did not require to be observed as in the other aspects as reliability and practicality

Validity in Language Testing

Validity is the most important characteristic of a test or assessment technique because it measures what it purports to measure. Without good validity, all else is lost. In the J. B. Heaton's book, *Writing English Language Test* (1998: 153), Heaton said that the validity of a test is the extent to which it measures what it is supposed to measure and nothing else. For example, if the teacher wants to measure the speaking skills, the test that is developed must be able to measure the ability to speak, not writing skills. Further, as insisted by Gronlund in Brown (2004: 22) the extent to which inferences made from assessment result are appropriate, meaningful, and useful in terms of the purpose of the assessment. There is no final absolute to measure the validity of a test established but several different kinds of evidence may be invoked in support. Arikunto (2006: 65) mentioned two types of validity. They are empirical validity and logical validity. Empirical validity is as same as quantitative analysis. It consists of concurrent validity

<https://assignbuster.com/definition-of-test-types-of-test-education-essay/>

and predictive validity whereas the logical validity consists of content validity and construct validity.

However, the writer only focused its investigating on content validation because this study dealt with logical validity that is a way of reviewing in rational by using descriptive analysis method and thus got involved with qualitative inquiries. Beside, because of the limitation in time and the other aspects in validity need the expert judges.

Content Validity

A test has content validity that relates the evidence to the content of the test. The content is related to the goal of what has been taught. Fulcher & Davidson (2007) defined content validity as any attempt to show that the content of the test is a representative sample from the domain that is to be tested. In the same line, Hughes (2003) said that a test is said to have content validity if its content constitutes a representative sample of language skills and structures, etc. with which meant to be concerned. For example, a grammar tests, it must be made up of items relating to the knowledge or control of grammar.

In relation to test's content validity in this study, test's content should be tested in accordance with the target competencies which reflected on Standar Kompetensi-Kompetensi Dasar in view of the school based curriculum to be achieved. These Standar Kompetensi-Kompetensi Dasar broken down into skills or behaviors that can be measured called indicator. In short, content validity is arranged based on the content in subject that

evaluated. Because those subjects are taught written in curriculum, it can be often called curricular validity.

According to John A. Upshur (1996: 63), content-related evidence of validity is assessed logically by carefully and systematically examining whether the method and content of the assessment procedure are representative of the kinds of language skill. It means that evidence of content-related validity requires the panel discussion as primary method to determine whether a test has content validity or not. In the panel discussion, the expert judges who considered having knowledgeable to do with the subjects tested. They asked to make judgments about the appropriateness of each item and overall coverage of the domain.

The School Based Curriculum

As cited in the National Education Standards (SNP) section 1 (15), School Based Curriculum is operational curriculum that developed and implemented by each educational unit. The developing of curriculum conducted by each educational unit has to follow based on the standard competence and the basic competence which is designed by Badan Standar Nasional Pendidikan (BSNP). It means that each educational unit is given the authorities in developing their curriculum. Its curriculum is designed as a reference in using the materials, teaching and learning techniques. Based on Regulation Minister of National Education no. 41/2007, The standard competence means minimum competency qualification for student which show mastery of knowledge, behavior, and skill of student which expected to be achieved in each level and/or each semester of certain subject whereas the basic competence means as a number of abilities that should be mastered by

<https://assignbuster.com/definition-of-test-types-of-test-education-essay/>

students in certain subject as a reference in developing indicator of competency.

Based on the school based curriculum, the standard competences and the basic competences of English in Vocational High School is divided into three different level which are level novice, level elementary, and level intermediate. In relation to this study, there is one Standard Competence and eight Basic Competences of English in Vocational High School for Level Novice. The Standard Competence is Berkomunikasi dengan Bahasa Inggris setara Level Novice. Then, the eight Basic Competences of English in Vocational High School for Level Novice are 1. 1 Memahami ungkapan-ungkapan dasar pada interaksi sosial untuk kepentingan kehidupan; 1. 2 Menyebutkan benda-benda, orang, ciri-ciri, waktu, hari, bulan, dan tahun; 1. 3 Mendeskripsikan benda-benda, orang, ciri-ciri, waktu, hari, bulan, dan tahun; 1. 4 Menghasilkan tuturan sederhana yang cukup untuk fungsi-fungsi dasar; 1. 5 Menjelaskan secara sederhana kegiatan yang sedang terjadi; 1. 6 Memahami memo dan menu sederhana, jadwal perjalanan kendaraan umum, dan rambu-rambu lalu lintas; 1. 7 Memahami kata-kata dan istilah asing serta kalimat sederhana berdasarkan rumus; and 1. 8 Menuliskan undangan sederhana.

The English subject itself put in the adaptive group that provides students the ability in the English communication in the context of material related to the majors' need either spoken or written. Also, English subject also directs students to be able in speaking in daily communication based on global requirement and developing communication at higher level. As stated in the curriculum that English subject aim for the learners to master basic

knowledge and skills to support the achievement of English language competency skills program and apply skills and mastery of English language skills to communicate both verbally and written at the intermediate level.

Items Test Analysis

The activity aims at increasing quality of a test called item test analysis.

Nitko (1996: 308) stated that item analysis refers to the process of collecting, summarizing, and using information about individual test items especially information about pupils response to items. Item test analysis is used to identify quality of the test whether it is good, low or not good. As stated by Anastasi and Urbina (1997: 184), the main aim of item analysis of the teacher made test is to identify its deficiencies. In conducting item test analysis, two methods are able to be used, namely qualitative and quantitative analysis. As stated by Suyata (2009: 16), items test can be analyzed by quantitative or empirical and qualitative or theoretical. Item test analysis in terms of qualitative covers the content of a test while difficulty level, discrimination power and distractors function involved quantitative analysis. In the same line, Purwanto (1992) said that by making analysis, it can be known the important things of every single item obtained, the extent to which difficulty level, whether item has discrimination power, and whether all alternative answer (options) attract the answer. The more explanation about difficulty level, discrimination power, and distractors function are as below:

Difficulty Level

Aiken (1994: 66) said that Difficulty Level is the opportunity correct answer a question at a certain skill level. It is usually expressed in the form of an index

<https://assignbuster.com/definition-of-test-types-of-test-education-essay/>

which has proportion range 0, 00 - 1, 00. If the item test has 0, 00 of Difficulty Level Index, it means that no students who answer correctly. While if the item test has 1, 00 of Difficulty Level Index, it means that student can answer correctly. In short, the higher of Difficulty Level Index, the easier of item test is understood and vice versa. Based on Crocker & Algina (1986) in calculating the item difficulty, divide the number of people answering the item correctly by the total number of people answering item. The proportion for the item is usually denoted as P and is called item difficulty. Arikunto (2006: 207) explained that a good test is about not too easy or too difficult. The easy test is not able to stimulate students learning. Contrary, the difficult test is able to make students desperate because of out of their reach.

As cited by Departemen Pendidikan Nasional, Direktorat Jendral Manajemen Pendidikan Dasar dan Menengah, Direktorat Pembinaan Sekolah Menengah Atas (2008) in Panduan Analisis Butir Soal,

The prediction information about the easy item test is as follow:

The distractor function doesn't work well,

Most students answer those item tests correctly; it means most students have understood the content in question.

The prediction information about the difficult item test is as follow:

Item tests maybe have mistake a key answer,

Item tests have two or more the correct answers,

<https://assignbuster.com/definition-of-test-types-of-test-education-essay/>

The content in question hasn't been taught yet or hasn't finished in its learning so that the minimum competency student have doesn't achieved yet,

Unsuitable measured content using format given in the item test,

Statement or item sentence is too complex and long.

Discrimination Power

Discrimination Power refers to measurement of the extent of the ability of items of achievement test to distinguish between students' high answers and students' low answers based on criteria. This notion is supported by Arikunto (2006: 211), Discrimination Power is item's ability to distinguish between students who are good and low capable. The proportion range of Discrimination Power Index is from -1, 00 to +1, 00. The higher of Discrimination Power Index, the more capable the item test distinguish between students who had already understood and hadn't already understood the content. An item test has negative Discrimination Power Index (<0) means it has more lower group answer correctly than upper group. The procedures process of discrimination index analysis, firstly score each student's test and rank order the test scores. Then, those scores are divided into three groups of students' score by 27% from the total number of students. They are upper group those who have high score, lower group those who have low score and middle group those who have middle score. Wiersma and Jurs (1990: 145) stated 27% is used because it has shown that this value will maximize differences in normal distribution while providing enough cases for analysis. Further, the number of people in the upper group

who answered the item correctly minus the number of people in the lower group who answered the item correctly, divided by number of people in the largest of the two groups.

As cited by Departemen Pendidikan Nasional, Direktorat Jendral Manajemen Pendidikan Dasar dan Menengah, Direktorat Pembinaan Sekolah Menengah Atas (2008) in Panduan Analisis Butir Soal,

The prediction information about disability item test to distinguish lower and upper student group is as follow:

Inappropriate key answer of item

Item tests have two or more the correct answers

Unclear of measured competency

The distractor function doesn't work well,

The content in question is too difficult, so many students are guessing.

Distractors Function

Distractor function is meant to know work or not the answer of item. Arikunto (2006: 220) stated a distractor function works well if it at least

selected 5% by test taker. This analysis is only used to analyze multiple choice item tests.

Theoretical Framework

In order to answer the research questions of this study, English items test analysis of Tes Kendali Mutu of first grade school year 2011/2012 at SMKN

<https://assignbuster.com/definition-of-test-types-of-test-education-essay/>

26 Jakarta is reviewed in terms of: (1) the relevance of English items test's content to the school based curriculum for level novice of Vocational High School; (2) validity; (3) difficulty level; (4) discrimination power; and (5) the distractor function. In relation to test's content validity in this study, qualitative analysis is used by analyzing items test in reference to the aspects of test validity. Measuring the validity of a test, the English items test of Tes Kendali Mutu must contain proper sample of relevant material in learning. The relevant material in learning can be found in the syllabus and curriculum for level novice of Vocational High School which is School-based Curriculum (KTSP). In order to answer whether English test of Tes Kendali Mutu has content validity, the writer needs Standar Kompetensi and Kompetensi Dasar in the school based curriculum to compare and match its items with the relevant syllabus and curriculum.

Furthermore, in order to support in answering the research question of this study, quantitative was conducted by using classical items test analysis through multiple choice test involving Difficulty Level, Discrimination Power, Distractor Function, and Validity. They are also considered to fulfill requirement criteria of quality a good test.