

**AN EVALUATION OF ENGLISH TEST
FOR SMK STUDENTS OF 2010
IN THE LOCAL EXAMINATION IN SAWAHLUNTO**

THESIS



ALFITRI YUSUF
NIM. 2009/51973

This thesis is submitted to fulfill one of the requirements
for the Degree of Magister Pendidikan (M.Pd)

**ENGLISH EDUCATION SECTION
LANGUAGE EDUCATION PROGRAM
GRADUATE PROGRAM
STATE UNIVERSITY OF PADANG
2011**

ABSTRAK

Alfitri Yusuf. 2011. Evaluasi Tes Bahasa Inggris Untuk Siswa SMK Tahun 2010 di Ujian Daerah Sawahlunto. Tesis: Padang: Program Pasca Sarjana Universitas Negeri Padang

Tes Bahasa Inggris untuk siswa SMK di Sawahlunto digunakan sebagai instrumen untuk mengukur sejauhmana siswa-siswa SMK di Sawahlunto dapat menguasai kompetensi dasar Bahasa Inggris yang ada dalam Kurikulum Tingkat Satuan Pendidikan SMK. Hasil tesnya digunakan sebagai bahan pertimbangan pengambilan keputusan dan penilaian tentang kemajuan siswa dan program perbaikan. Kenyataannya, tes tersebut tidak diujicoba dan dianalisis sebelum diujikan kepada peserta tesnya sehingga tidak ada informasi tentang kualitas tes tersebut. Beberapa masalah berkaitan dengan aspek aspek tes muncul ketika tes itu diujikan. Hal ini menyebabkan guru-guru Bahasa Inggris SMK di Sawahlunto sulit mengetahui hasil belajar siswa mereka yang sebenarnya berdasarkan hasil tes tersebut. Penelitian ini merupakan sebuah penelitian evaluasi. Penelitian ini bertujuan untuk mengevaluasi tingkat kesulitan, daya pembeda, fungsi pengecoh, reliabilitas, dan validitas dari tes Bahasa Inggris SMK Kelas XI pada ujian daerah tahun 2010 di Sawahlunto. Penulis telah menganalisis tes tersebut menggunakan program ANATEST versi 4.0.3 untuk memperoleh data empiris tentang karakteristik soal. Hasil analisis menunjukkan kelemahan tes. Dalam hal tingkat kesulitan soalnya, ditemukan 52 % soal terkategori mudah. Dalam hal daya pembeda soalnya, 90 % soal terkategori buruk. Dalam hal fungsi pengecohnya, 90.7 % pengecoh dalam soal tes tersebut tidak berfungsi. Reliabilitas tes terkategori rendah karena ditemukan indek koefisien reliabilitasnya adalah 0.25. Tes tersebut valid dalam hal isi tesnya. Namun, tes tersebut mempunyai kelemahan dalam hal kesesuaian indikator soal yang ada dalam kisi-kisi tesnya dengan soal-soal tesnya. Kelemahan juga ditemukan dalam hal tampilan tesnya. Berdasarkan temuan-temuan tersebut, disimpulkan bahwa tes tersebut belum bagus. Tes tersebut masih memerlukan perbaikan.

ABSTRACT

Alfitri Yusuf. 2011. An Evaluation of English Test for SMK Students of 2010 in the Local Examination in Sawahlunto. Thesis: Padang: Graduate Program State University of Padang

English test for SMK students in Sawahlunto is used as an instrument to measure how well the students can master basic competences in the school-based curriculum. The test result is used as inputs to make decisions and judgments about the students' progress and remedial programs. In fact, the test was not tried out and not analyzed before being given to the test takers so that there was no information about the test quality. Some problems related to the aspects of test occurred on the day of test. This made English teachers in Sawahlunto difficult to know their real students' achievement based on the test result. This was an evaluation research. The purpose of the research was to evaluate the difficulty indices, item discriminations, function of distractors, reliability and validity of the English test for SMK of Grade XI in the local examination of 2010 in Sawahlunto. The writer analyzed the English test by using ANATEST program version 4.0.3 to get the empirical data about the characteristic of each item. The result of the analysis showed the weaknesses of the test. In terms of the difficulty indices, it was found 52 % items were in the level of criteria "easy". In terms of the discrimination indices, 90 % items were in the level of criteria "poor". In terms of the function of distractors, 90.7 % distractors in the test items did not function. The test reliability was in the level of criteria "low" because it was found the coefficient index was 0.25. The test was valid in the content because all basic competences which were taught have been covered in the test items. However, the test has the weaknesses in terms of the suitability of item indicators in the test blueprints with the test items. The weaknesses were also found in terms of the test appearance. Based on the findings, it is concluded that the test is not good yet. It still needs improvements.

CHAPTER I

INTRODUCTION

A. Background of the Problem

Education Law No. 20/2003 verse 58 article 1 requires teachers to evaluate their students' achievement in order to get information about students' progress. This means that evaluation toward students' achievement is one of the teachers' duties. The evaluation is necessary to do in order to give feedback to schools particularly teachers for continual and better improvement related to either the teaching programs or the teaching and learning processes. The evaluation is also important as the actualization of schools and teachers' responsibility to ensure whether the teaching programs run well or not in achieving the instructional objectives. The results of evaluation can be used as inputs to improve the quality of teaching programs for students' benefits.

Related to this, test can be used as an instrument of evaluation. As a matter of fact, test has been used by educational institutions, schools, and teachers to measure how far the students can achieve the instructional objectives of the teaching and learning processes in the classrooms. In other words, test is needed to provide information about the achievement of students. The result of test is often used as inputs for schools especially teachers to make decisions for the remedial programs to improve the students' achievement.

Test as an instrument of evaluation should be valid and reliable. Validity and reliability of a test are very essential. They will contribute to the truth

of information which is usually used as inputs for making decisions and judgments by not only stake holders in education but also teachers at schools. Moreover, a good test will provide real information and data about not only students' achievement and teachers' practice in teaching but also schools' progress in implementing the contents of curriculum. In contrast, a bad test will provide unreal information and data about students' achievement, teachers' practice in teaching, and schools' progress. It can create an incorrect decision making and judgment for improving the quality of teaching programs.

In the practice, test as an instrument of evaluation can be summative test. In Indonesia, it is well known as "Ujian Semester". It actually has been used at schools in Indonesia in order to get information about their students' achievement at the end of semester. Theoritically, summative test is a test designed to measure students' achievement at the end of semester after all materials have been taught. The test result is usually used by teachers as inputs to make decisions whether the students need remedial programs or not and whether the students can go to the next grade or not. It is also used by schools as inputs to see how far the instructional objectives have been achieved by the students and how well the teachers can implement their teaching programs to achieve the instructional objectives stated in the school-based curriculum, well known as "Kurikulum Tingkat Satuan Pendidikan" (KTSP). Moreover, the test results are also used as inputs to evaluate the teaching programs at schools including teachers' teaching techniques, lesson plans, media, and teaching materials. In the practice, low

achievements usually indicate that there may be some weaknesses or problems related to the teaching programs and teaching techniques conducted by teachers. Based on the information, teachers can do revisions or improvements which are useful to improve students' achievement.

In relation to this, the local government of Sawahlunto through the educational department has conducted a summative test, well known as “Ujian Standar Kota” (a local examination in Sawahlunto) held at the end of semester at schools in Sawahlunto. The test is important to get information and data about the schools' success in implementing the school-based curriculum. It is also used as an instrument to evaluate the quality of schools every semester. By this method, the government wants to see the strengths and weaknesses of teaching programs at schools in Sawahlunto. All decisions and judgments of educational programs for schools and teachers are based on the information.

The examination is held at schools by using a multiple-choice test. This test is designed by some selected teachers from every subject e.g. Biology, Mathematics, Religion, Bahasa Indonesia, IPA, IPS, and English. They are selected and appointed by the educational department (Diknas) to design a test in which the test items are constructed related to their subjects and grade levels (Grade I to VI for SD, Grade VII to IX for SMP, Grade X to XII for SMA and SMK) based on the school-based curriculum. All items are made on the basis of indicators in the blue prints of the test. The test items which have been constructed

by test designers are usually saved by Diknas in the item banks. The test items are usually retested for the next year examination.

Based on the writer's observation, the local test was designed by the selected teachers from SD, SMP, SMA, and SMK of semester I of 2010/2011 academic year in Sawahlunto. They were senior English teachers in Sawahlunto. They were selected by Diknas because they were considered having experiences in designing tests. They worked in groups in Diknas. In the groups, the test designers discussed and did the test development such as constructing the test specification e.g. the blue prints based on standard competences (SK), basic competences (KD), indicators, and number of items, determining the test technique e.g. multiple-choice test, preparing the test construction e.g. materials, questions, options, language, and the key answers. Finally, the test items were printed out, copied and distributed by Diknas to the schools on the day of test.

On the day of test, the test items were given to the groups of test takers in the classrooms. Each group consisted of 20 SMK students whose the level of ability was high and low. Each group of test takers was watched by two teachers. The teachers were in charge of distributing the test sheets and the answer sheets to the test takers, explaining the test instructions, watching the test takers from cheating, collecting the test takers' answer sheets, and submitting them to the test administrators or the committee.

Through an interview with the test designers of English for SMK, the writer assumed that the test designers know how to design an English test and

how to construct the test items. They were also assumed to know about the standard quality of a good language test such as test validity, reliability, practicality, and the test wash back.

However, when the test items were tested to the students in the local examination of semester I of 2010/2011 academic year in December 2010, the writer found some problems on the test items. Based on the writer's observation during the test, it was found that pictures displayed on the test papers were not clear, for example the picture number 1, 2, and 3. Some options used in the test items were ambiguous because they seemed to have more than one correct answer, for example the options in the item no. 23, 43, 44, 49, and 50. Item no. 33 did not exist in the test sheet. These made the test takers disappointed and underestimated the test. They felt they couldn't do the test well although they prepared themselves well before the test. They also felt anxious that they would get low score on their academic record. These also made the SMK English teachers in Sawahlunto especially the writer himself unable to see the real students' achievement based on the test. It was difficult to get the real information about the students' mastery toward the materials taught by the teachers in the classroom based on the test.

Ideally, a language test which will be used as an instrument to measure students' achievement should be good. To ensure a language test is good or not, an item analysis can be applied. Item analysis involves the calculation of difficulty indices, discrimination indices, and function of distractors of the test items in the

case of multiple-choice test as stated by Hughes (2002, p. 225). The difficulty indices, discrimination indices, function of distractors as well as reliability and validity of a test will contribute to the quality of a test. Therefore, the analysis of difficulty indices, discrimination indices, function of distractors, reliability and validity of the English test is very important for test designers to do before giving the test to the test takers. This technique can be used to help test designers to find out the good and bad items. The results of the analysis can be used by test designers as inputs to maintain the good items and to revise the bad items for better test construction.

Since the English test is usually used as an instrument of evaluation for making decisions and judgments toward students' achievement, teachers' work and schools' progress in Sawahlunto, the test is necessary to be analyzed. It is important to find out the weaknesses of the English test for the test improvement before being used as an instrument of evaluation. Hence, the writer analyzed the English test for SMK used in the local examination in Sawahlunto.

B. Identification of the Problem

Conceptually, the English test used in the local examination should be designed by considering some aspects including test development, test technique, test administration, reliability, validity, and the wash back. Ignoring these aspects in designing the test can affect the quality of the test itself. This can make the test not function well as an instrument for gathering information about the students' achievement as inputs for continual improvement.

In the practice, the writer found some problems related to some aspects of the English test. They consist of the aspect of test administration, test development, test reliability, test validity.

The problems appeared because no analysis was applied to the test before being given to the test takers so that there was no information about the quality of the English test. Analysis on the test is important to do to ensure whether the test has fulfilled the criteria of a good test. It is also important to do to find out the weaknesses of the test in order to be revised. The revision is necessary to improve the quality of English test so that the test results can be used as inputs to make decisions and judgments about the students' progress and the success of teaching programs. This can be done by analyzing the difficulty indices, item discriminations, function of distractors, reliability and validity of the English test.

C. Limitation of the Problem

Based on the identification of the problem, the researcher limits the research on the evaluation of English test for SMK students of Grade XI in the local examination of semester I of 2010/2011 academic year in Sawahlunto. The research was focused to know the difficulty indices, item discriminations, function of distractors, reliability and validity of the English test. These aspects would be analyzed through the item analysis. In the analysis, these aspects were judged by using the criteria suggested by experts.

D. Statement of the Problem

Based on the limitation of the problem, statement of the problem of the research was as follows: “How are the difficulty indices, item discriminations, function of distractors, reliability and validity of the English Test for SMK students of Grade XI in the local examination of semester I of 2010/2011 academic year in Sawahlunto?”

E. Research Question

Based on the statement of the problem, the research questions were made as follows:

1. How are the difficulty indices, item discriminations, and function of distractors of the English test for SMK students of Grade XI in the local examination of semester I of 2010/2011 academic year in Sawahlunto?
2. How is the reliability of the English test for SMK students of Grade XI in the local examination of semester I of 2010/2011 academic year in Sawahlunto?
3. How is the validity of the English test for SMK students of Grade XI in the local examination of semester I of 2010/2011 academic year in Sawahlunto?

F. Purpose of the Research

Based on the research questions, the purpose of the research is:

1. To know the difficulty indices, item discriminations, function of distractors of the English test for SMK students of Grade XI in the local examination of semester I of 2010/2011 academic year in Sawahlunto.
2. To know the reliability of English test for SMK students of Grade XI in the local examination of semester I of 2010/2011 academic year in Sawahlunto.
3. To know the validity of English test for SMK students of Grade XI in the local examination of semester I of 2010/2011 academic year in Sawahlunto.

G. Significance of the Research

1. For teachers:
 - a) The teachers can find out the weaknesses of multiple-choice test used in the English test in the local examination of Sawahlunto as inputs for better test construction.
 - b) The teachers can get ideas how to analyze an English test in the form of multiple-choice test used to measure their students' achievement.
2. For schools:
 - a) The schools can obtain the real data about the strengths or weaknesses of English test used as the instrument to measure the students' achievement.

- b) The schools can obtain the real data of the students' achievement as inputs to evaluate the teaching programs.
- 3. For the development of practical theories:
 - a) The theories and ideas in the research can be useful for the development of practical theories .
 - b) The research findings can be useful for other researchers for further research.

H. Definition of the Key Terms

1. The English Test

The English test is a test given to the SMK students of Grade XI of semester I of 2010/2011 academic year in Sawahlunto in order to measure the students' achievement toward the instructional courses of English taught by their teachers.

2. The Local Examination

The local examination is a kind of summative assessment using an English test for SMK students in the form of multiple-choice test conducted in order to evaluate the schools' progress and the students' achievement at semester I of 2010/2011 academic year in Sawahlunto.

CHAPTER V

CONCLUSION, IMPLICATION, AND SUGGESTION

A. Conclusion

The English test for SMK students of Grade XI of 2010/2011 in the local examination in Sawahlunto was used as an instrument to measure how well the students can master basic competences in the school-based curriculum. The test was designed in the form of multiple-choice test. It was tested at the end of semester simultaneously at SMK in Sawahlunto. The test results were used as inputs to make decisions and judgments about the students' progress and remedial programs. In the practice, the test was not analyzed before being given to the test takers so that there was no information about the quality of the test.

After analyzing the English test and getting the results, the writer concludes that:

1. The difficulty indices of the English test are in the level of criteria “easy” as it was found 52 % of the test items could be answered easily. This happened because of providing clues in the stems and using keywords in the stems as the key answers. In terms of the discrimination indices, the discrimination indices of the English test are in the level of criteria “poor” as it was found 90 % of the test items could not discriminate between the high and low students. This happened because most distractors did not function well. Concerning the function of distractors, most distractors do not function as it was found 90.7 %

of the distractors in the test items do not function. This happened because of using keywords in the stems as the key answers.

2. The test reliability is low as it was found the index of reliability coefficient is 0.25. This happened because of the easy items, the poor item discrimination, and the distractors which did not function.
3. In terms of the test validity, the test can be judged “valid” in the content as it was found all basic competences which were taught have been covered in the test items. However, the test still has weaknesses in terms of the suitability of item indicators in the test blueprints with the test items, and in terms of the test appearance.

As a whole, it can be concluded that the English test is not good yet. It still needs improvements. Hence, the English test should be revised.

B. Implication

The research findings bring some practical implications to the test development. They are as follows:

1. In the case of multiple-choice test, test should be designed by considering about the characteristics of a good test including the difficulty indices, discrimination indices, function of distractors, reliability, and validity.
2. The content area which is intended to measure should be written in the test blueprints in which the basic competences, item indicators, materials, and number of items are stated clearly.
3. Test items should be suitable with item indicators in the test blueprints.

C. Suggestion

Based on the research findings, the writer wants to provide some suggestions, as follows:

1. English teachers should try out, analyze, and revise the tests which they design before being tested to their students. It is important to do to ensure that the tests are reliable and valid as an instrument to measure the students' achievement.
2. Education Department (Diknas) and schools in Sawahlunto need to invite the practitioners of evaluation or language testing to train the English teachers about how to design a good test. It is because designing test is one of the teachers' duties at schools. English teachers usually evaluate their students' progress by using tests. Besides, designing a good language test is not easy. Hence, well training and knowledge about test design are necessary for English teachers.
3. English Teachers Forum or "Musyawarah Guru Mata Pelajaran" (MGMP) in Sawahlunto should be more active in preparing and discussing tests that will be used by English teachers. This can be a good way to share knowledge and experience of making good items between senior and junior English teachers in Sawahlunto.
4. It is expected that other researchers will conduct further research on the evaluation of test and test design.

REFERENCES

- Arikunto, S. 2009. *Dasar-Dasar Evaluasi Pendidikan* (Edisi Revisi). Jakarta: Bumi Aksara.
- Ary, D. L. C. Jacobs and A. Razavieh. 2002. *Introduction to Research in Education* (6th ed.). Belmont: Thomson Learning, Inc.
- Bachman, L F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Best, J and J. Khan. 2003. *Research in Education* (9th ed.). Boston: Pearson Education Company.
- Brown, A. 2001. *Research Report: Item Performance and Reliability of the English Language Proficiency Test of applicants for immigration to Canada*. <http://www.Englishraven.com/pdf>. (July 26th, 2011)
- Brown, H. D. 2004. *Language Assessment: Principle and Classroom Practice*. San Fransisco: Longman.
- Brown, J. D and T. Hudson. 2002. *Criterion-Referenced Language Testing*. Cambridge: Cambridge University Press.
- Burton, S. J, R. R. Sudweeks, P.F. Merrill and B. Wood. 1991. *How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty*. Brigham: Brigham Young University Testing Services.
- Chan, S. 2010. *Research Note: Demonstrating Cognitive Validity and Face Validity of Writing Items Summarize Written Text and Write Essay*. <http://www.pearsonpte.com/research/pdf>. (July 26th, 2011)
- Diaz, Lynne and T. Rico. 2004. *Teaching English Learners: Strategies and Methods*. Boston: Pearson Education, Inc.
- Farhady, H. 1986. *Characteristics of Language Tests: Total Test Characteristics*. <http://www.socialresearchmethods.net/kb/relanval.php>. (June 29th, 2011)
- Finocchiaro, M and S. Sako. 1983. *Foreign Language Testing: A Practical Approach*. New York: Regents Publishing Company, Inc.

- Gay, L.R and P. Airasian. 2003. *Educational Research: Competencies for Analysis and Application* (7th ed.). New Jersey: Pearson Education, Inc.
- Henning, G. 1987. *A Guide to Language Testing: Development-Evaluation-Research*. Massachusetts: Newbury House Publishers.
- Hatch and Lazaraton. 1991. *The Research Manual: Design and Statistics for Applied Linguistics*. Los Angeles: Newbury House Publishers.
- Hadley, A. O. 2001. *Teaching Language in Context* (3rd ed.). Houston: Thomson Learning, Inc.
- Hughes, A. 2002. *Testing for Language Teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Harmer, J. 2007. *The Practice of English Language Teaching* (4th ed.). London: Pearson Education, Inc.
- Mertler, C. A and C. M. Charles. 2005. *Introduction to Educational Research* (5th ed.). New York: Pearson Education, Inc.
- O'Loughlin, K. 2009. Research Summary: Investigating the lexical validity of Test of English Academic for Foreign Language Students. <http://www.pearsonpte.com/research/documents>. (July 27th, 2011)
- Pae, H. 2007. *Differential Item Functioning and Unidimensionality in the Pearson Test of English Academic*. <http://www.pearsonpte.com/research/pdf>. (July 27th, 2011)
- Popham, J. W. 1981. *Modern Educational Measurement*. Los Angeles: Prentice Hall, Inc.
- Richards, J. C. 2001. *Curriculum Development in Language Teaching*. Cambridge: Cambridge University Press.
- Richard, P and L. Amato. 2003. *Making It Happen: From Interactive to Participatory Language Teaching-Theory and Practice* (3rd ed.). New York: Pearson Education, Inc.
- Weir, C. J. 1990. *Communicative Language Testing*. London: Prentice Hall International (UK) Ltd.

Weir, C. J. 1993. *Understanding and Developing Language Tests*. London: Prentice Hall International (UK) Ltd.

----- 2010. *Juknis Analisis Butir Soal*. Jakarta: Direktorat Pembinaan SMA.