

**AN ANALYSIS OF ENGLISH FINAL TEST OF GRADE XI
AT MAN-s IN WEST SUMATERA**

Thesis

*Submitted As Partial of Fulfillment of Requirements to Obtain Strata One
(S.1) Degree in English Department of State University of Padang*



**Iswahyudi
83596/2007**

Advisors:

- 1. Dr. Desmawati Radjab, M.Pd**
- 2. HavidArdi, S.Pd, M.Hum**

**ENGLISH DEPARTMENT
FACULTY OF LANGUAGE AND ART
STATE UNIVERSITY OF PADANG
2012**

HALAMAN PERSETUJUAN

Judul : An Analysis of English Final Test of Grade XI
at MAN-s in West Sumatera

Nama : Iswahyudi

NIM/BP : 83596/2007

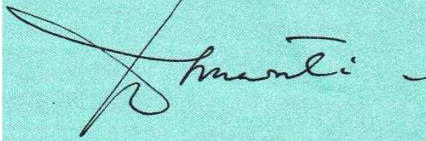
Program Studi : Pendidikan Bahasa Inggris

Jurusan : Bahasa dan Sastra Inggris

Fakultas : Bahasa dan Seni

Disetujui oleh:

Dosen Pembimbing I



Dr. Desmawati Radjab, M.Pd.
NIP:19501231 197703 2 002

Dosen Pembimbing II



Havid Ardi, S.Pd., M.Hum
NIP:19790103 200312 1 002

Diketahui,
Ketua Jurusan Bahasa dan Sastra Inggris



Dr. Kurnia Ningsih, M.A.
NIP: 19540626 198203 2 001

HALAMAN PENGESAHAN LULUS UJIAN SKRIPSI

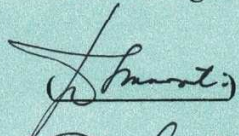
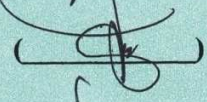
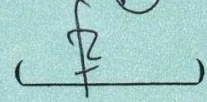
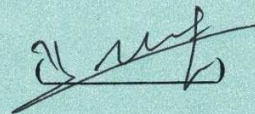
**Dinyatakan Lulus Setelah Dipertahankan di Depan Tim Penguji Skripsi
Jurusan Bahasa dan Sastra Inggris
Fakultas Bahasa dan Seni
Universitas Negeri Padang**

**An Analysis of English Final Test of Grade XI
at MAN-s in West Sumatera**

**Nama : Iswahyudi
NIM/BP : 83596/2007
Program Studi : Pendidikan Bahasa Inggris
Jurusan : Bahasa dan Sastra Inggris
Fakultas : Bahasa dan Seni**

Padang, 28 Mei 2012

Tim Penguji

	Nama	Tanda Tangan
1. Ketua	: Dr. Desmawati Radjab, M.Pd.	
2. Sekretaris	: Havid Ardi, S.Pd., M.Hum.	
3. Anggota	: Dra. Hj. Yenni Rozimela, M.Ed., Ph.D.	
4. Anggota	: Dr. Kusni, M.Pd.	
5. Anggota	: Yuli Tiarina, M.Pd.	

ABSTRAK

ISWAHYUDI.2012.*An Analysis of English Final Test of Grade XI at MAN-s in West Sumatera.* (Unpublished Thesis). Padang: FBS- Universitas Negeri Padang

Pembimbing : 1. Dr. Desmawati Radjab, M.Pd

2. Havid Ardi, S.Pd, M.Hum

Skripsi ini merupakan analisa soal Bahasa Inggris yang digunakan oleh MAN di Sumatera Barat pada tahun ajaran 2010/2011. Topik yang dibahas dalam analisa ini adalah kualitas soal, relevansi soal terhadap kurikulum Bahasa Inggris, dan bagaimana soal bahasa inggris itu dibuat. Permasalahannya adalah siswa mengalami kesulitan dalam menjawab dan adanya ketidaksesuaian kompetensi yang diukur pada soal Bahasa inggris pada dua semester sebelumnya.

Metode yang dipakai dalam penelitian ini adalah metode deskriptif karena penelitian ini bertujuan untuk menggambarkan kualitas soal Bahasa Inggris, mengidentifikasi kesesuaian soal terhadap kompetensi yang diukur, dan mengetahui bagaimana soal bahasa inggris dibuat. Objek penelitian ini adalah soal Bahasa Inggris yang digunakan MAN di Sumatera Barat pada tahun ajaran 2010/2011. Sedangkan, Sumber data untuk penganalisaan ini adalah lembar jawaban siswa-siswa dari MAN Lembah Melintang, MAN Air Bangis, MAN Sumpur, dan MAN 3 Padang. Sedangkan metode analisa penelitian ini menggunakan metode *item analysis* untuk mengukur kevaliditasan, reliabilitas, tingkat kesukaran, daya pembeda, dan keefektifan pengecoh setiap butir soal.

Hasil temuan penelitian ini adalah soal-soal Ujian Akhir Semester kualitas kurang baik. Dilihat dari segivaliditas, hanya 5 butir soal yang valid. Dari segi reliabilitas, soalBahasaInggristersebut tidak reliabel walaupun memiliki nilai reliabilitas tinggi. Dilihat dari tingkat kesukaran soal, 7 (14%) soal mudah, soal sedang 40 (80%) soal dan 3 (6%) soal sulit bagi siswa-siswa tersebut. Untuk Daya Pembeda Soal, soal yang termasuk dalam kategori soal jelek 14 (28%) soal, soal cukup 26 (52%) soal, dan 10 (20%) soal baik. Dari segi keefektifitas pengecoh, 11 (22%) soal yang distraktornya berfungsi, 39 (78%) soal yang distraktornya membutuhkan revisi dan pengantian dengan pilihan yang baru. Relevansi soal ke kompetensi yang diukur, hanya satu soal yang tidak mengukur kompetensi yang telah diajarkan pada semester tersebut. Sedangkan soal Bahasa Inggris tersebut dibuat oleh guru Bahasa Inggris yang dipilih dari pengalaman yang berasal dari wilayah 2 dan diedit oleh wilayah 3, dan soal ini tanpa adanya *try out* sebelum dilaksanakan keseluruh siswa MAN di Sumatera Barat.

ACKNOWLEDGEMENTS

Alhamulillahirobilalamin, Praise be to Allah SWT that this thesis has been completed. Then, *Shalawat* and *Salam* are sent to the beloved Muhammad SAW Prophet who has struggled in creating the world be better.

The writer's deepest gratitude is addressed to Mrs. Dr. Desmawati Radjab, M.Pd and Mr. Havid Ardi, S.Pd, M.Hum, for their thought, guidance, motivation, support, patience, constructive commands, and available time given to the writer during the completion of this thesis.

The writer presents his sincere appreciation to his academic advisor, Mr. Amri Isyam, M.Pd, for his motivation in completing this research. My gratitude is sent to Mrs. Yenni Rozimela, M.Ed, Ph.D, Mr. Dr. Kusni, M.Pd, and Yuli Tiarina, S.Pd, M.Pd, as the reviewers of this research. Thank you very much for all of the English Department lectures of State University of Padang who have taught and guided the writer in his study and his thesis. Then, thank you so much for chairman of English Department, Mrs. Dr.Hj. Kurnia Ningsih, M.A, her secretary, Mrs. Dra. An Fauzia Rozani Syafe'i, M.A, and also to her staffs, Dean of Faculty of Language and Art of State University of Padang, Mr. Prof. Dr. H. M. Zaim, M.Hum and his staffs have helped the researcher in managing all formal needs during this thesis arrangement.

The researcher's special gratitude goes to his beloved parents, Junaidi (the late), and Erlisna for their love, their care, and their pray. This is also dedicated to his beloved sisters and his beloved bothers, who always support him in completing his study. The writer also wishes to express his deepest appreciation to his relative family in Padang, Nenek, Gaek (the late), Da Eri, Pak Iman, Pak Iwan, Pak Faisal, BukWita, BukWinda, and his lovely younger brothers and sister who have cared, motivated, and prayed during his study. I dedicated this thesis for you all.

The researcher greatest thanks go to all of his friends, especially for Mike and Fanny, Kepend-1 2007, all 2007 English Students, and all seniors of English Department. Greatest grateful is sent to his hometown close and lovely friends for support, great friendship, and share idea.

May Allah SWT the almighty richly bless the people and institution that help him all the ways.

Padang, June 2012

The Writer

TABLE OF CONTENTS

Abstrak.....	i
Acknowledgements	ii
Table of Contents	iii

CHAPTER I INTRODUCTION

A. Background of the Problem	1
B. Identification of the Problem	4
C. Limitation of the Problem	5
D. Formulation of the Problem	5
E. Research Question	5
F. Purpose of the Research	6
G. Significance of the Research	7
H. Definition of the Key Term	7

CHAPTER II REVIEW TO RELATED LITERATURE

A. Test	9
1. Type of Test	10
a. Proficiency Test	11
b. Diagnostic Test	11
c. Placement Test	11
d. Achievement Test	12
2. Steps in Constructing Test	13
3. Quality of Good Test	16
a. Validity	18
1) Content Validity	18
2) Construct Validity	19
3) Face Validity	21
b. Reliability	22
1) Stability	23
2) Equivalent	23

3) Split Half Method	24
c. Item Difficulty index.....	26
d. Item Discrimination index	27
e. Distractors	28
B. Teaching English Subject at Senior High School	29
1. Objective of English subject at Senior High School	30
2. Scope of English Subject at Senior High School	30
C. English Final Test	32
1. Definition	32
2. Function of English Final Test	33
3. Characteristic of English Final test	34
D. Review of Related Studies	35
E. Conceptual Framework	37

CHPATER III RESEARCH METHODOLOGY

A. Research Design	39
B. Research Object	39
C. Source of Data	40
D. Technique of Collecting Data	40
E. Data Analysis	41
1. Quality of Good Test	41
a. Validity Index	42
b. Reliability Index	43
c. Item Difficulty Index	45
d. Item Discrimination Index	47
e. Effectiveness Distractors	45
2. Relevant to Competences in Curriculum	49

CHAPTER IV FINDINGS AND DISCUSSION

A. Data Description.....	50
B. Analysis and Finding	50

1. Validity	50
a. Item Validity	50
b. Content validity	52
c. Construct Validity	53
d. Face Validity	54
2. Reliability	56
3. Item Difficulty	57
4. Item Descriminability	60
5. Effectiveness of Distractors	62
6. The Procedure of Developing Test	64
C. Discussions	68

CHAPTER V CONCLUSIONS AND SUGGESTIONS

A. Conclusions	75
B. Suggestions	76

REFERENCES

CHAPTER 1

INTRODUCTION

A. Background of the Problem

In classroom, there are three general activities that are always performed by a teacher. They are planning, implementing, and evaluating. These activities cannot be separated because they are an interlinked system that is to gain the objectives of learning that have been stated at the beginning of all semesters. The evaluating, for example, is done to know effectiveness of teaching process of a subject. It is explained more by Richard (2001) the evaluation given by teacher is to check how the subject runs. Anderson and Morgan (2008) add that a teacher uses the evaluation to find contextual information factors that may affect the aim of learning in particular curriculum aspect. Therefore, the evaluation can be served as important information that would or would not improve the teaching learning process.

Similarly, the three activities are also involved in teaching English as one of subjects in Indonesia. Evaluation is used to check students' English capability, to describe the achievement level of the students, to calculate the students' mark, and to report the students' result to the students' parents at the end of each semester. Moreover, in English subject, DEPDIKNAS (2007:5) states that the most important purpose of the evaluation is to control the English subject process. Consequently, all

high school students are always evaluated through English evaluation tool at the end of each semester. This evaluation tool is commonly known in Indonesian as a test.

There are many types of test that can measure students' ability in English, for example, multiple choice test. The test consists of item questions, a key and some distractors. The question should be answered by choosing the correct answer from the distractors. In the teaching learning English context, the items test must have good quality in order to measure the students' real capability in English, to describe the students' achievement, and to match the students' competences with competences stated in English curriculum area at the beginning of the English program. According to Brown (2004:19) the good test has characteristics, they are: a) Validity, b) Reliability, c) Practicality, d) Authenticity, and e) Washback. The characteristics above, the test will be appropriate to measure accurately what will be to measure. In addition, Anderson and Morgan (2008) add that the good items of the test have good materials: 1) good item difficulty, 2) good item discriminability, and 3) good distractors. All these characteristics are important in a test in order to gain the purpose of learning.

In applying the English final test, all senior high schools hold it to measure the students' English capability. As senior high schools, all of MAN-s in West Sumatera which are Islamic based high school hold the English test to measure students' ability in English. The English test, which is held in Islamic senior high schools, is different from other senior high school. The Islamic schools tests are designed by the

Department of Religious Affair (Depag). The department designs the test since the first semester at first grade until the first semester at third grade, while Final Examination (UN) is designed by the Department of Education. Similarly, they construct all the tests based on school based curriculum (KTSP 2006) areas.

In preliminary research, a brief observation was held on January 31st, 2011, the writer interviewed an English teacher at Madrasah AliyahNegeri (MAN) LembahMelintang, the teacher explained that some of his students took English test in the two last semesters and did not really understand the reading texts on the test sheet. It was not only caused by lack in English vocabulary but also it was caused by the limited time to understand more the texts. He also said that some items of the test did not measure the stated competences on English curriculum.

At the same time, the writer did a survey at Madrasah AliyahNegeri (MAN) LembahMelintang and he asked the second grade students to response to questionnaire (see Appendix A) sheets as well as the questionnaire direction about the test in last semester in order to know what are their problems on the test. And then, he collected questionnaire from 61 students in two classes after ten minutes. Based on the information, he found that 67.21 % of students got difficulties in understanding some long texts on test sheet in short time test. It was found 63.93% of students said that it did not cover all materials that had been given by their English teacher. 49.18% of students mentioned that the test used many difficult words and idioms. 34.43% of students argued that many options on some items were almost a like the options seem

all correct answer. They also said that there are items that had long stem of question, so they did not know what the question exactly asked for.

Based on the problems and the survey above, it is needed to identify and investigate the truth of the problems. Because of that, the writer wants to do the research on the English test for second grade of MAN-s in West Sumatera at 2010/2011 academic year. The writer did an analysis of quality of the English test with information from the students' responses toward the test items, which was given by the second grade students of MAN-s in West Sumatera. The test has been held in the last final semester test, particularly at Wednesday on 8th June 2011. The English final test consists of 50 items, and each item has 4 distractors and an option as a key.

B. Identification of the Problem

Based on the survey to the second grade students in MAN Lembah Melintang, it was found that they had problems in answering the question of the item test. As the teacher mentioned, some item test did not cover all competences on school based curriculum (KTSP 2006). In addition, as stated before the students had problems in comprehending of questions based on text and found difficult words on some of the items test. These conditions are not only caused by the students' inability in English, but also it can be caused by the test itself.

Related to test analysis, there are many relevant research can be done, such as studying characteristics of good test, the quality of the test item, the test materials,

grammatical questions of items test, level of vocabulary used in the test, test accommodations, and relevancy of the test to curriculum.

C. Limitation of the Problem

Based on the explanation above and considering the writer's ability, knowledge, time, and skill, he only focused on finding the value for each characteristic of good test, the quality of items test, and the relevancy of the test to curriculum.

D. Formulation of the Problem

Related to the discussion above, this study is formulated as follow:

“How is quality of the English final test on 2010/2011 academic year of second grade of MAN-s in West Sumatera based on item analysis?”

E. Research Questions

In order to make this research to be applicable and easy to be conducted, the formulation of problem of this research is expanded to be several research questions as follow:

1. How is validity of the English final test of second grade of MAN-s in West Sumatera at 2010/2011 academic year?
2. How is reliability of English final test for second grade of MAN-s in West Sumatera at 2010/2011 academic year?

3. How is the item difficulty of the English final test of second grade of MAN-s in West Sumatera at 2010/2011 academic year?
4. How is item discriminability of the English final test of second grade of MAN-s in West Sumatera at 2010/2011 academic year?
5. Does the English final test for second grade of MAN-s in West Sumatera at 2010/2011 academic year provide effective distractors?
6. How is the English Final test of second grade of MAN-s in West Sumatera at 2010/2011 developed?

F. Purpose of the Research

The purpose of this research is to identify whether the test has good quality. It could be described as below:

1. To investigate the validity of the English final test.
2. To investigate the reliability of the English final test.
3. To identify the value of the item difficulty index of the English final test.
4. To identify the value of the item discrimination index of the English final test.
5. To identify the effectiveness the distractors of the English final test.
6. To investigate how the English final test developed.

By doing so, the test designers know the quality of the test and how the items work. Then, they would make revises and improvement to be better test if it is needed. For example, which item should be changed is. As a result, qualified test measure the real students' achievement. In addition, the English teachers gather the

information as feedback from the students on the English course. The information will be useful in selecting or creating teaching material and strategies, and in reexamining the effectiveness the teaching learning process in order to be more effective teaching learning process in the next teaching period.

G. Significance of the Research Purpose

This research is hoped to give significant input to the test designer about English subjetin order to make better English items test. It is hopefully served a part as guidance to create new test in order to be more appropriate to measure the English competences for the next English final test.

H. Definition of Key Term

- | | |
|------------------|--|
| 1. Test | : an examination of students' knowledge
consists of question for them to answer. |
| 2. Item test | : a single number of test with individual score |
| 3. Test analysis | : studying of test in order to understand more
about quality of the test item |
| 4. Item analysis | : studying of the quality of the items in order to
reveal how difficult the item test is, how well
the discrimination the item is, and how effective
distractors used by the item is. |
| 5. Test Validity | : extent to which the test measures what it
claims to measure |

- 6. Test Reliability : test produces the consistency of scores in a different time
- 7. Item difficulty : items of a test have difficulty level for the students
- 8. Item discriminability : items that differentiate between high ability students and low ability students.
- 9. Distractors : incorrect option in the items test
- 10. Final test : a test is to measure whether student pass a level in order to enter to the higher level

CHAPTER V

CONCLUSIONS AND SUGGESTIONS

A. Conclusions

Based on data analysis of the English final test of MAN-s in West Sumatera at 2010/2011 academic year, and its discussion, it could be concluded that only five items had good quality to measure the students' ability, while the rest test items need improvement, and revision even change to new item. Even though, this research foundas follows:

1. The English final test had generally invalid items. It might be caused by the stem of the items, and ineffective distractors.
2. English final test is unreliable to measure students' competence.
3. The average of item difficulty of English final test was medium item.
4. The average of item discriminability of English final test was satisfactory to discriminate between high and lower students' ability.
5. There were many ineffective distractors of English final test items.
6. The English final test was not tried out and item analysis. The test was made by English teachers and team designers from Bukit Tinggi, Padang Panjang, Agam, and Pasaman.

B. Suggestions

Since the KTSP is used as guidance in teaching learning affects the teaching learning process even the designing of tool evaluation (test). The quality of the test depends on the teachers' ability in English and teachers' experiences in creating test. Based on research finding, there were test items which would need improvement in order to create more qualified test. Therefore, there are several suggestions suggested for as following:

1. Teacher

Since the teachers are as the first sources in creating the test, it is suggested that the teachers create test based on KTSP that have been taught to their students. Besides that, the teachers should analyze the given test in order to know how the English subject runs.

2. Test designers

Test designers are suggested to administrate the trial test firstly, analyze the trial result and final result test, and design the test by guiding to KTSP and Passing Standard of competence in order to ensure the quality of giving test, ensure the stem of test items are written grammatical, and highlight negative and/or double negative on the items' stems by underlying and capitalizing. Besides that, since it was found that there were several wrong key items used by the teachers to check their students' answer sheets from the four MAN-s in this research, it is also suggested

the test designers to provide the keyanswer for the tests in order to keep same perception in using the keys test items.

3. Next Researchers

For next researcher who wish to conduct deeper research in this research, it is expected to seek other issues that had been not covered by the researcher such measuring the amount time in doing the test to the total test items, and investigating the vocabularies level and material by providing on the English test.

REFERENCES

- Anderson, P and G. Morgan. 2008. *Developing tests and questionnaires for a national assessment of educational achievement*. Washington DC: The World Bank.
- Arikunto, Suharsimi. 2009. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Brown, F. G. 1983. *Principles of Educational and Psychological Testing*. 3rd Edition. New York: Holt, Rineheart, and Winston.
- Brown, H.D. 2004. *Language Assessment: Principles and Practices*. New York: Pearson Education, Inc.
- Brown, J. D. 1995. *The elements of language curriculum*. Boston: Heinle & Heinle Publisher.
- Gay, L.R., and P. Airasian. 2001. *Educational Research: Competency Analysis and Application Sixth Edition*. New Jersey: Prentice-Hall inc.
- Hughes, H. 2003. *Testing For Language Teacher*. Cambridge: Cambridge University Press
- Iskandar. 2008. *Metodologi Penelitian Pendidikan dan Sosial*. Jakarta: Gaung Persada Press.
- Izard, J. 2005. *Overview of Test Construction*. Modul 6. Retrieved on January 2012 from <http://www.sacmeq.org> or <http://www.unesco.org/iiep>.
- Kelly T. L. 1939. *The Selection of Upper and lower groups for Validation of test items*. Retrieved on January 2012: <http://paraonline.com/files/upperandlower.pdf>
- Marion, Scott F., and Jim Pellegrino. 2009. *Validity Framework For Evaluating The Technical Quality Of Alternate Assessments Based On Alternate Achievement Standards*. Chicago: National Center for the Improvement of Educational Assessment (NCME).
- Martinez, R. J., M. Rafael, I. Martín, and M. Eva Trigo. 2009. *Evaluation of five guidelines for option development in multiple-choice item-writing*. *Psicothema*, Vol. 21, No. 2. Retrieved on June 2011:

<http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=72711654024>

- McAlpin, M., and Carol, H. 2001. *Chapter 4: New Assessment Methods*. The Online Tutoring Skills Project. Retrieved on June, 2011: <http://otis.scotcit.ac.uk/onlinebook/>
- Mitra, N.K., Nagaraja, H.S., Ponnudurai, G. and Judson, J. P. 2009. *The Levels Of Difficulty and Discrimination Indices In Type A Multiple Choice Questions Of Pre-clinical Semester 1 Multidisciplinary Summative Tests*. Kuala Lumpur: Academic Affairs Department of International Medical University.
- Morales, R.A. 2009. *Evaluation of Mathematic Achievement Test: A Comparison between CTT and IRT*. Northern Samar: Time Taylor Academic Journals.
- Muslich, M. 2008. *KTPS: Dasar Pemahaman dan Pengembangan*. Jakarta: Bumi Aksara
- Nazir, M. 2005. *Metode Penelitian*. Bogor: Ghalia Indonesia.
- Ricahard, J.C. 2001. *Curriculum Development in Language Teaching*. Cambridge: Cambridge University Press.
- Richard, R., and T. Rodgers. 2001. *Approaches and Methods in The Language Teaching*. New York: Cambridge University Press
- Suryosubroto. 1990. *Prinsip Kurikulum dalam Pendidikan*. Jakarta: Duta Wahana University Press.
- Young, V.M., and D.H. Kim. 2010. *Using Assessment for Instructional Improvement: a literature review*. Educational Policy Analysis Archives, Vol. 8, No. 19. Retrieved on June, 2011: <http://eppa.asu.edu/ojs/article/view/809>
- Zimmaro. Dawn M. 2004. *Writing Good Multiple-Choice Exams*. Measurement and Evaluation Center. Retrieved on June, 2011: <http://www.utexas.edu/academic/mec>
- Badan Standar Nasional Pendidikan (BSNP). 2007. *Standar Penilaian Pendidikan*. Peraturan Menteri Pendidikan Nasional Republik Indonesia. Jakarta.

BadanStandarNasionalPendidikan (BSNP). 2008. *Kurikulum SMA/MA*. DepartementPendidikanNasional. Jakarta.

Israel Academy of Science and Human, Ministry of Education, Culture, and Sport, and Rothschild Foundation (YadHanadiv). 2005. *A Framework For Curricula: academic Studies and development*. The Knowledge Base for Assessment and Evaluation in Education.Retrievedon June, 2011: <http://academy.ac.il/science> policy.

DepartemenPendidikanNasional. 2003. *Kurrikulum 2004: StandarKompetensi(BahasaInggrisSekolahMenengahAtasdan Madrasah Aliyah*. Jakarta: DepartemenPendidikanNasional.

DepartemenPendidikanNasional. 2007. *StandarPenilaianPendidikan*.PeraturanMenteriPendidianNasionalRe publik Indonesia.Jakarta.

DepartemenPendidikanNasional. 2009. *BukuSaku:Kurikulum Tingkat SatuanPendidikan (KTSP) (sekolahMenengahAtas)*. Jakarat: DepartemenPendidikanNasional.

Professional Testing Inc. 2006.*Conduct the Item Analysis*.Conduct the Item Analysis. Building High Quality Examination Programs.Retrieved on 7th June, 2011: http://www.proftesting.com /test_topics/steps_9.shtml