



HOW ETHIOPIAN STANDARDIZED NATIONAL EXAMINATIONS ACHIEVE THEIR GOAL? 2014/15 UNIVERSITY ENTRANCE EXAMINATION EXAM IN FOCUS

Hadya Hassen

M.A in TEFL, Department of Foreign Languages and Literature, Addis Ababa University, Addis Ababa, Ethiopia

Corresponding Email: hadyahassen9@gmail.com; Tel: +(251)912634491

Cite this article:

Hadya Hassen (2022), How Ethiopian Standardized National Examinations Achieve their Goal? 2014/15 University Entrance Examination Exam in Focus. African Journal of Social Sciences and Humanities Research 5(3), 1-14. DOI: 10.52589/AJSSHR-PBV0DRVO.

Manuscript History

Received: 16 April 2021

Accepted: 5 May 2021

Published: 7 July 2022

Copyright © 2022 The Author(s).

This is an Open Access article distributed under the terms of Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), which permits anyone to share, use, reproduce and redistribute in any medium, provided the original author and source are credited.

ABSTRACT: *The main purpose of this study was to examine the effectiveness of the items of the 2014/15 university entrance examination of the English language based on the parameters of facility value and discrimination index. The study has used a descriptive research design concerned with the quantitative methods which reveal the statistical result of the exam. The data were collected from the database of the National Educational Assessment and Examination Agency (NEAEA). The study covered all of the 120 items of the Ethiopian University Entrance Examination (EUEE). One-fourth of the examinees' i.e. 40,400 examinees' results were taken for analysis. Secondary data was also used in the analysis process. The collected data were analyzed quantitatively. Within the two main parameters of item analysis, namely, item facility value and discrimination index of the study; the items were categorized into four ranges based on their effectiveness. Findings of the study revealed that 33 items were very difficult, and 27 items were not discriminating at all. The findings revealed that high numbers of the EUEE examination items were not functioning as expected, which directs many items to be either discarded or improved. This meant the yearly implementation of post-test item analysis is required for better preparation of functional standardized exam preparation for the next periods.*

KEYWORDS: Achievement test, multiple-choice test items, Item analysis, Item facility value, Discrimination index.



INTRODUCTION

Background

“A test in simple terms, is a method of measuring a person’s ability, knowledge or performance in a given domain” (H.D.Brown 2003). Harold S. Madsen (1983) asserted that testing is an important part of every teaching and learning experience. The definition sounds fairly simple, but in fact, constructing a good test is a complex task involving both science and art Brown (2003). According to Brown, most language tests measure one's ability to perform language; to speak, write, read, or listen to a subset of language. Carroll (1968) as cited in Bachman (1995) defined test as” 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.” (Carroll 1968: 46)

There are different categories of language tests according to their purpose. Bailey (1998) noted eight kinds of language tests have different kinds of purposes of language tests, language aptitude tests, language dominance tests, proficiency tests, admission tests, placement tests, diagnostic tests, progress tests, and achievement tests. However, James Dean Brown (1996) categorized language tests into two main categories; norm-referenced and criterion-referenced and grouped all the above kinds of tests in these categories. An achievement test is grouped in criterion-referenced language tests.

One of the achievement tests given in Ethiopia is the EFL university entrance examination which contains 120 multiple choice test items. All the language skills and sub skills are considered to be included. The time allotted to accomplish the exam is 1 hour and 30 minutes which means one minute is given for each item. Harrison (1983) cited in Kassaw Baye (2006) explained that an achievement test looks back over a long period of time and is intended to show the standard which the students have now reached in relation to other students at the same stage.

High stake tests in Ethiopia are all standardized and designed in multiple-choice questions and students take federally organized exams including EFL in addition to other subjects. The purpose of these entrance exams is the selection for joining higher institutions. These exams are conducted by the national educational assessment and examination agency (NEAEA) and marked electronically. To this end, the research is concerned with item analysis of university entrance examination of the year 2005 E.C through data obtained from NEAEA.

In the context of practical realities, these national-level standardized EFL examinations are creating fear and anxiety for those who take them. Hence the form, content, and style of these exams are left similar since years ago, teachers in these grade levels spend their time collecting previous sheets and discussing their answers during the time of the lesson.

Indeed, some researchers researched related aspects. Among them, Kassaw (2006), in his thesis, examining the content validity of UEEs, the multiple-choice items do not function properly for all skills and sub-skills. In addition, unfamiliar task design and test format lead candidates to anxiety. Consequently, even high achiever students may not perform as they are expected. As a result, detailed research is needed to be conducted on the analysis of each item. Simachew (2012), indicated the washback effect of UEEs. And Kassaw(2006) and Kifle (1995) again investigated the content validity of the national examination.



However, as there is a lack of post-test item analysis practice by NEAEA and no concern is given for post-test item analysis, UEEs are always similar in form and content through the curriculum, and the teaching-learning materials are changed or improved in the Ethiopian situation. Therefore, the researcher intended to assess how the items are functioning.

Objective of the Research

In addition to the general matters of test characteristics validity and reliability Madsen,(1983) referred to be concerned on the effect of examinations by taking time to evaluate individual items by using the ways we can improve our tests. According to J, Alderson (1995), a post-test analysis should be made on different aspects and facility value and discrimination index are mentioned for this purpose.

The main purpose of this study is to find out the effectiveness of EFL UEE test items to see how they function for the targeted population. The effectiveness of a test is determined by the effectiveness of each item, and each EFL test item stands for a particular skill or sub-skill intended to be assessed.

- Assessing how difficult each EFL test items are
- Investigate to what extent each item of the EFL exam discriminates between high achieving and low achieving students.

Research Design

This study reveals descriptive statistical analysis results of test items. “Item-analysis is usually done for two purposes; one for selecting the “best” items that will remain on a revised and improved version of the test. The other is simply to investigate how well the items on a test are working with a particular group of students.” (McNamara 2000). Brown (1971) mentioned that item analysis has two purposes: Firstly, it enables the identification of defective items, to improve the test and evaluation procedures. Secondly, through indicating which items or material students have and have not mastered. This research intends to analyze teachers’ need to evaluate each item individually by using the two common item analysis parameters (Item Facility and Discrimination index). Hence, it reveals the numerical results of the items the study is uses quantitative technique.

Data Collection and Analysis

One national-level standardized EFL test with 120 MCQ items was chosen and the recorded scores of the examinees were taken from the database of ENEAEA. The exam was prepared for grade 12 students in 2005 E.C. with the purpose of selection for joining higher institutions. This examination is set by subject area experts at the national level. It is a paper-pencil test yearly administered by ENEAEA. Moreover, this exam as an EFL test consists of seven sections in order to address language skills and sub-skills. These are word order, paragraph coherence, vocabulary, grammar, reading, communicative activities, and writing.

170,000 examinees have taken the examination. Taking the four codes as a clue for separation code, 14 examinees, 40,400 (¼) examinees’ results were selected using a simple random sampling technique.



The collected data had gone through SPSS and excel spreadsheet.

All 120 items for comprehensives were taken for analysis. In doing term facility value, the proportion of students who answered the questions correctly were taken as a formula ($p=C/TN$). Here, the level of difficulty was classified into four groups based on the criteria of Bailey (1998).

1. above 0.85 =very easy
2. below 0.3 =very difficult
3. 0.3-0.39 =reasonably acceptable and
4. 0.4-0.85 =ideal items

Although there are various similar ways of calculating the discrimination power, the researcher has used the simplified technique; taking the upper 27% (10,908) and lower 27% (10,908)of the examinees' results. In this work the researcher used the following formula to calculate the discrimination index of the items:

$$\frac{\text{The upper group of students-lower group of students}}{\text{Total number of students}} = \frac{(U_s-L_s)}{T_s}$$

The items were categorized into four groups by the level of their discriminating power based on the range given by Ebels (1992).

1. If $DI \geq 0.40$, the item is functioning quite satisfactorily.
2. If $0.30 \leq DI \leq 0.39$, little or no revision is required.
3. If $0.20 \leq DI \leq 0.29$, the item is marginal and needs revision.
4. If $DI \leq 0.19$, the items should be eliminated or completely revised.

RESULTS AND DISCUSSIONS

Multiple-choice tests are of considerably widespread use as a means of objective measurement. The main reason behind such popularity is the many dominant advantages associated with multiple-choice tests. They can be used for diagnostic as well as formative purposes and can assess a broad range of knowledge. In addition, they are scored easily, quickly, and objectively either by human beings or by scoring machines. These and many similar advantages make multiple-choice tests suitable for a wide range of purposes ranging from classroom achievement testing to large-scale standardized tests. Thus, improving the quality of multiple-choice test items appears to be of a lot of importance, (Baghaei & Amrahi 2011).

120 multiple-choice items were coded based on their category in accordance with the booklet and analyzed based on their order. Therefore, the general display of the analysis of the items is put in the table. Items facility value and discrimination index are rather displayed in graphs.



It is not as easy to design a question to assess a student's language ability like vocabulary definition, reciting grammatical rules, reading comprehension, writing skill, and answering the questions after listening to a particular speech. According to Madsen (1983), if a language test is designed properly, it has different advantages such as elimination of foreign language anxiety and it allows learners to think of they can accomplish any task in the target language.

One best multiple choice question (MCQ), if properly written and well constructed, is one of the strategies of the assessment tool that quickly assesses any level of cognition according to Bloom's taxonomy. Therefore, it is important for teachers to evaluate their MCQ items to see how effective they are in assessing the knowledge of their students, and in predicting their total test scores. Difficulty and discrimination indices are fundamental tools to check whether the MCQs are well constructed or not. Brown(2003) also stated issues as he called five cardinal criteria "for testing a test"; which are test practicality, reliability, validity, authenticity, and washback. Every individual criterion also includes different things to be assessed specifically. Looking at an item's difficulty and discrimination will assist the test developer in determining what is wrong with individual items. Item and test analysis provide empirical data about how individual items and whole tests are performing in real test situations.

Very easy questions may not sufficiently challenge the ablest student. However, having a few relatively easy questions in a test may be important to verify the mastery of some course objectives. Very difficult questions, if they form most of a test, may produce frustration among students. Some very difficult questions are needed to challenge the best students. When tests are too easy or too difficult, the scoring distribution will tend to unnaturally concentrate at one end of the continuum. As a result, it is difficult to distinguish candidates' abilities at the concentrated end. This, inevitably, also results in loss of person separability or reliability (Henning, 1987).

Table 1: The general display of the analysis result of all item

ITEM NO	IF	DI	ITEM NO	IF	DI
1 word order	.58	-.1	61	.27	.53
2	.78	.22	62	.38	.21
3	.59	.20	63	.42	-.1
4	.38	-.5	64	.38	.35
5	.83	.21	65	.39	.06
6 paragraph coherence	.26	.24	66	.29	.21
7	.81	.23	67	.46	.40
8	.68	.31	68	.55	.71
9	.39	.21	69	.23	.48
10	.42	.27	70	.36	.21
11	.51	.40	71	.35	.21
12	.30	.50	72	.42	.27
13 Reading	.41	.33	73	.24	-.3
14	.47	.21	74	.16	.33
15	.54	.43	75	.41	.04
16	.17	.36	76	.14	.02
17	.50	.45	77	.30	.05



18	.50	.07	78	.56	.29
19	.26	.20	79	.44	-.3
20	.22	-.3	80	.37	-.1
21	.20	-.1	81	.18	.0
22	.37	.42	82	.56	.23
23	.16	.01	83	.40	.24
24	.34	.21	84	.33	.27
25	.27	-.2	85	.52	.51
26	.60	.21	86	.42	.61
27 vocabulary	.22	.21	87	.40	.16
28	.36	.45	88	.47	.60
29	.42	.41	89	.55	.52
30	.03	.21	90	.53	.41
31	.28	.61	91	.31	.15
32	.16	.43	92	.57	.40
33	.51	.40	93	.17	.22
34	.28	.41	94	.75	.0
35	.55	.36	95	.42	.32
36	.08	.90	96	.43	.0
37	.15	.04	97	.63	.16
38	.19	.36	98	.47	.72
39	.47	.25	99	.50	.41
40	.36	.22	100	.73	.31
41	.50	.08	101	.53	.23
42.	.20	.06	102	.77	.30
43	.24	.04	103	.61	.29
44	.32	.34	104	.45	.27
45	.39	.21	105	.26	-.3
46	.53	.42	106 writing	.53	.34
47	.40	.23	107	.45	.30
48	.45	.27	108	.51	-.1
49	.42	.21	109	.46	.11
50	.39	.01	110	.64	.20
51	.32	.33	111	.26	.35
52	.29	.06	112	.45	.44
53	.51	.22	113	.13	.21
54	.37	.57	114	.32	.20
55	.48	.69	115	.24	.35
56	.46	.21	116	.20	.26
57	.47	.6	117	.27	.21
58	.52	.50	118	.33	.35
59	.53	.34	119	.31	.44
60	.39	.21	120	.28	-.1

**Table2-The list, number, and percent of all items in terms of facility value**

	Item facility		
	Item no	Total	%
Very easy	no very easy item		
Ideal	1,2,3,5,7,8,10,11,,13,14,15,17,18, 26,28 ,29 ,33,35, 39, 40, 41, 46, 47, 48, 49, 53, 55, 56, 57,58,59,63,67, 68, 72,75,78,79,82,83,85,86,87,88,89,90,92,94,95,96,97,98,99,100,101,102,103,104,106,107,108,109,110,112	64	53.3
Reasonably acceptable	4,9,12,22,24,44,45,50,51,54,60,62,64,65,70,71,77,80,84,91,114,118,119	23	19.16
Very difficult	6,16,19,20,21,23,25,27,30,31,32,34,36,37,38,42,43,52,61,66,69,73,74,76,81,93,105,111,113,115,116,117,120	33	27.5

Table 3-the list, number, and percent of total items in terms of discrimination index

	Discrimination index		
	Item no	Total	%
Quite satisfactory	11,12,15,17,22,28,29,31,32,,33,34,46,54,55,57,58,61,67,68,69, 85,86,88,89,90,92,98,99,112	29	24.16
No or little revision required	2,6,7,8,10,13,14,16,30,35,38,39,40,44,47,48,51,53,59,62,64,66 ,72,74,78,82,83,84,93,95,100,101,102,103,104,106,107,110,111 1113,115,116,118	43	35.3
Marginal(needs revision)	3,,5,9,24,26,2,45,49,56,60,65,70,71,73,87,91,97,,109,114,117, 119,	21	17.5
Should be eliminated	1,4,18,19,20,21,23,25,36,37,41,42,43,50,52,63,75,76,77,79,80, 81,94 96,105,108,120	27	22.5

**Table 4 –item facility result of items in their category**

Total items	Very easy	Ideal	Reasonably acceptable	Very difficult
Word order(5)	-	4	1	-
Paragraph coherence (7)	-	4	2	1
Reading (14)	-	6	2	6
Vocabulary(31)	-	15	5	11
Grammar(28)	-	12	9	7
Communicative activities(20)	-	17	1	2
Writing(15)	-	6	3	6

Regarding facility value, in this study, many items 33 (27.5%) are very difficult. As English is a common course and the exam has to be given for both social and natural science students, we need to take into account that there are different levels of students from different social groups. But based on the result of this study, it is difficult to say such things are considered.

When we see the proportion of difficulty level of items from the seven sections of the language section in the exam, vocabulary items are the most difficult items; 11 (9.2%) of the items range below 3.0 (extremely difficult). However, grammar items are better than vocabulary items; 7 (5.8%) of the items are difficult and need to be improved or discarded. Both reading and writing items are proportionally easier than vocabulary and grammar. 6(5%) items from writing items and 6(5%) others from reading are found to need to be improved or discarded since they are very difficult questions. However, there is no very difficult item in the word order, but 1 item from paragraph coherence and 2 items from communicative activities are found to be very difficult items. Most of the items are found to be in the range of ideal and reasonably acceptable facility value.

The average result of the difficulty level of the items is 0.3 which reveals that the exam is not as extremely difficult in terms of facility value. Rather, it is in the reasonably acceptable range which needs some revision. In addition to the average result of facility value, 53 % (more than half) of the items are not very difficult.

Discriminating power is one powerful indicator of item effectiveness. If the test and an item measure the same ability or competence especially when an exam is a group of high stake tests, we would expect that those items that have a high overall test scorer would have a high probability of being able to answer the item. We would also expect the opposite, which is to say that those having low test scores would have a low probability of answering the item correctly. Thus, a good item should discriminate between those who score high on the test and those who score low. The discrimination index is a useful measure of item quality whenever the purpose of a test is to produce a spread of scores, reflecting differences in student achievement, so that some distinctions may be made among the performances of respondents, (Hotiu 2006). Item discrimination (ID) indicates the degree to which an item separates the students who performed well from those who performed poorly.



Regarding the discrimination index, in the present study, 27 (22.5%) of the total items have an index of discrimination < 0.1 (should be illuminated). This shows that no focus is given to the effectiveness of the individual items, Experience of time-wise test item analysis helps to avoid this kind of a high number of not functioning test items.

Proportionally from the seven parts of the examination, most of the test items in the vocabulary section have a good discrimination power. Among 31 vocabulary items, 10 are in the range of satisfactorily discriminating and another 10 items need no or little revision. The grammar test items have also a good power of discrimination. From a total of 28 items, 6 items are quite satisfactory and 10 items need no or some revision. Except for one of the items in word order, all of them do not discriminate at all. The other parts of test items are found to be not discriminating, and they need to be discarded or improved in order to increase the effectiveness of the exam.

The average discrimination index of the whole test is 0.14 which means most of the items need revision. This indicates that most of the items in the exam do not have a good range of discrimination index.

The correlation of items facility value and discrimination index is negative, which means that items with good facility value are not likely to be discriminating, and items that have a satisfactory index of discrimination were found to be very difficult. Only 5 reading, 3 vocabularies, 2 dialogue, and 1 writing items are problematic both in difficulty and facility value.

The quality of test items may be further improved based on an action that can be taken in reviewing the distractors by the item writer based on the calculated discrimination and difficulty index values. Items showing poor discrimination should be referred back to the content experts for revision to improve the standard of these test items. It is important to evaluate the test items to see how effective they are in assessing the knowledge of the students based on the difficulty and discrimination indices of the test items.

The difficulty of the instruction can be a cause for the ineffectiveness of the items. In addition to the numerical result of the items, in this study, inconsistent language and unclear instruction are observed in the exam. For example, the instruction for the grammar section lacks consistency.

It reads: Questions 58-85 are incomplete sentences. There are four alternative words or phrases, A-D, Given below each question. Choose the word that best completes the sentence and blacken the letter of your choice on the separate answer sheet provided.

Some items do not meet their objective. The mismatch between the instruction and the intended objective to meet within the item is one factor that makes the item ineffective. To make the item effective there should be only one objective consistent with the instruction given. But when we see some items in the examination, there are items that don't examine the intended ability of students. For instance, in the last section of the exam, the writing skill of students is the intended skill to assess. But most of the items under this category are items that miss their objective. For example:



Item number 110 has written as below

➤ *Not suitable for children under 10! This is most likely a warning:*

A. *in a movie*

C. *in an x-ray room*

B. *on a box of whisky*

D. *on a pack of cigarettes*

The above question is prepared to assess students' writing skills, but it is not likely to assess the intended skill.

Distractor efficiency is one basic thing in the effectiveness of the item. All distractors must be functional. If distractors are not plausible enough, the effectiveness of the item could be damaged. As a result, it affects the quality of the exam. In this exam, there are almost no functioning distractors which were chosen by less than 5% of the examinees.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

In this study the researcher has analyzed all the items of the national university entrance examination of 2005 E.C, which has seven sections; word order, vocabulary, grammar, reading, paragraph coherence, communicative activities, and writing. As a result, based on the above discussion, the following conclusions have been drawn.

According to the result of the study, 1 item from paragraph coherence, 6 items from reading, 11 items from vocabulary, 7 items from grammar, 2 items from communicative activities, and 6 items from the writing section of the exam were found to be very difficult. Totally 33 (27.5 %) of the items in the exam are very difficult based on the range of facility value. Therefore, from the above data, one can conclude that not functioning items are too much according to the standardization level of the exam. These items do not fall in the standard range of facility value, which leads to an incorrect generalization about students' results.

According to the discrimination index, 2 word order items, 6 reading items, 7 vocabulary items, 7 grammar items, 3 communicative activity items, and 2 writing items were exhibited not to be discriminating (should be eliminated). Totally, 27 (22,5%) of the exam items are under the standard range based on the discrimination index. So, based on this result, it is possible to say that the quality of many items in the exam is under the standard range of the discrimination index.

As can be seen from Tables 4 and 5, the most difficult items are from the vocabulary part of the exam. The grammar part of the exam also has a high number of difficult items. The reading items are also found to be poor in difficulty level and discrimination power. 6 items were found to be very difficult. Unlike the above four sections, the other three sections: word order, paragraph coherence, and communicative activities are not difficult and are discriminating.

In this study, the correlation of facility value and discrimination index of items were negative, which means that all of the difficult items were not similarly bad in discriminating between high and low achiever students. And items that have a bad discrimination index are not the



most difficult. Of problematic items, only 10 are both difficult and they are better in discriminating in comparison with others.

Recommendations

High stake tests have a great role in identifying students' learning backgrounds, deciding their future educational careers, and have a great role in creating either anxiety or positive feeling towards a subject. So test makers need to be careful in preparing this kind of test.

Therefore, based on the conclusions, the following recommendations have been forwarded.

- Once the test is administered, its effectiveness has to be assessed based on different criteria, like facility value, discrimination index, and distractor efficiency in order to decide about the score of students. Therefore, the researcher would like to stress that the examination agency needs to develop the habit of doing item analysis in order to assess the practical quality of every year's examination before scoring the exam.
- Very difficult and very easy items need to be properly reconstructed and revalidated.
- As not functioning items affect the quality of the exam, it leads students, teachers, and other stakeholders to generalize incorrectly about their students; therefore it is better if not functioning items are improved or discarded before scoring the university entrance examination.
- This high stake test has great relevance for curriculum designing and materials preparation. It is better if the ministry of education develops a habit of having detailed reports of the exam rather than generalized, pass /fail reports.
- Based on the result of the study, vocabulary items were found to be the most difficult part of the exam. If the vocabulary is somewhat difficult, the item will likely measure reading ability in addition to the achievement of the objective for which the item was written. Therefore, in preparing for the exam the researcher recommends that it is better to use difficult and technical vocabulary only when essential for measuring the objective.
- As distractors have a great role in ensuring the quality of the item, making the distractors plausible enough and functional so that they will be chosen by students who score higher in the whole test result is essential. As a result, they will have positive discrimination power.
- The difficulty of instruction affects the effectiveness of the items. In the examination instructions with not clear and inconsistent language are observed. So, instructions must be set in clear, short, and precise language as much as possible.
- After the exam is administered, it needs to be checked to see if the item is perfect in assessing only one objective or not since an item is expected to assess only one objective. Doing this exam would be better if the irrelevant redundancy is removed from it. For instance, the last section of the exam (writing part) contains different items with different objectives rather than assessing writing skills. So it needs to be improved



and reconstructed to directly assess the intended student's ability. Rather, it leads the student into confusion over what he or she is being asked to do.

REFERENCES

- Alderson, J. C, Clapham, C and Wall, D, (1995), *Language Test construction And Evaluation*. Cambridge: Cambridge University Press.
- Bachman, L.F. (1990), *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L.F (1996), *Language Testing in Practice*. Oxford, Oxford University Press.
- Baily, Katherine, M,(1998). *Learning About Language Assessment*, Heinle Heinle Publishers.
- Bloom, B.S (1956) *Taxonomy of Educational Objectives*. Handbook 1: The cognitive Domain New York.
- Boophathiraj. C(2013) *Analysis of test items on difficulty level and discrimination index in the test for research in education*. International journal of social science & interdisciplinary research. Vol 2(2) online available at Indian research journal. com. Retrieved on March 20/2014.
- Brown, H.D, (1996), *Testing in Language Programs*. Prentice-Hall Regents.
- Crocker. L. & Algina, J, (1986) *Introduction to Classical And Modern Test Theory*. New York: Holt, Rinehart, And Winston.
- Cyril.Weir (2005), *Language Testing and Validation*. For Research in Testing Evaluation and Curriculum. Roehampton University Press.
- Davidson, F.and Brian K.Lynch (2002) *Testcraft: A Teacher's Guide To Writing And Using Language Test Specification*. Yale University Press
- Ebel RL.(1972) *Essentials of Educational Measurement* (1st Ed) New Jersey: Prentice-Hall.
- Ebel. L.E & Frisible, D.A.(1991) *Essentials Of Educational Measurement*. Prentice-Hall.
- El---Uri FJ, Malas N, *Analysis of use of a single best answer format in an undergraduate medical examination* .*Qatar Medical Journal* 2013;1http:// dx.doi.org/10.5339/ qmj.2013.1. Retrieved in march14/ 2014
- Harrold S.Madsen.(1983) *Techniques in testing*.Oxford University Press.
- Henning (1987) *A Guide To Language Testing Development, Evaluation, Research*. London: Newbury House Publisher.
- Hotiu, A,(2006), *the relationship between item difficulty and discrimination indices in multiple-choice tests in a physical science course*. M.A thesis, Florida Atlantic University.
- Kassaw baye (2006) *An exploration of content validity of 1997 E.C EHEFC English Language Examination*. M.A Thesis .unpublished.
- Madsen, S, (1967) *English Language Testing in Ethiopia: The ESLCE Examination*, Ethiopian Journal of Education.
- Madziah and Lide,(2006), *classical and Rasch analysis of dichotomously scored reading comprehension test items*. Malaysian Journal of ELT research. vol.2, International Islamic University Malaysia.
- Mein Shih, Y(2010) *An item analysis of test items on difficulty level and*



- discrimination index in the test for research in education.*
- McNamara, T,(2000), *Language Testing*, Oxford: Oxford University Press.
- Messick, S.(1989), *Validity: Educational Measurement (3rded)* New York, American Council & Macmillan.
- Omirin(2007) *Difficulty and discrimination indices of three multiple-choice tests using the confidence scoring procedure.* Educational research and review Vol.1.Academic journal. Online At<http://www.Academicjournals.org/ERR>. Retrieved on February 21, 2014.
- Rahim and Jaleel,(2002), *Analysis of one best MSQ:the difficulty index, discrimination index, and Distractor efficiency:* journal of Pakistan medical association.
- Simachew Gashaye(2011) washback of the University Entrance English Examination(UEEE) on teachers and students practices: the case of preparatory school in Amhara National Region.
- Wood, D.A,(1960), *Test Construction Development and Interpretation of Achievement Test*, Columbus, OH. Merrill Books



APPENDIXES

INSTITUTE OF EDUCATIONAL RESEARCH (IER)

ADDIS ABABA UNIVERSITY (AAU)

ETHIOPIAN UNIVERSITY ENTRANCE EXAMINATION (EUUEE)

ENGLISH, GINBOT 2005/JUNE 2013

BOOKLET CODE: 14

SUBJECT CODE: 01

TIME ALLOWED: 2 HOURS

GENERAL DIRECTIONS

THIS BOOKLET CONTAINS ***ENGLISH*** EXAMINATIONS. THE CODE FOR THIS EXAMINATION IS 01 AND THE CODE FOR THIS PARTICULAR BOOKLET IS **14**. PLEASE COPY THESE CODES ON YOUR ANSWER SHEET WHERE READS ***BOOKLET CODE AND SUBJECT CODE***. AND BLACKEN THESE CORRESPONDING BOXES IN THE COLUMNS BELOW EACH NUMBER.

IN THIS EXAMINATION, THERE ARE A TOTAL OF **120 QUESTIONS** DIVIDED INTO **SEVEN SECTIONS**. EACH SECTION CONTAINS **MULTIPLE CHOICE QUESTIONS** CONSISTING OF FOUR POSSIBLE ANSWERS. CAREFULLY SELECT THE BEST ANSWER AND BLACKEN ONLY THE LETTER OF YOUR CHOICE ON THE SEPARATE ANSWER SHEET PROVIDED. FOLLOW THE INSTRUCTIONS ON THE ANSWER SHEET AND THE EXAMINATION PAPER CAREFULLY. USE ONLY **PENCIL** TO MARK YOUR ANSWERS YOUR ANSWER MARK SHOULD BE HEAVY AND DARK, COVERING THE ANSWER SPACE COMPLETELY. PLEASE ERASE ALL UNNECESSARY MARKS COMPLETELY FROM YOUR ANSWER SHEET.

YOU ARE ALLOWED TO WORK ON THE EXAM FOR 2 HOURS. WHEN TIME IS CALLED, YOU MUST IMMEDIATELY STOP WORKING, PUT YOUR PENCIL DOWN, AND WAIT FOR FURTHER INSTRUCTION.

ANY FORM OF CHEATING OR AN ATTEMPT TO CHEAT IN THE EXAMINATION WILL RESULT IN AN AUTOMATIC DISMISSAL FROM THE EXAMINATION HALL AND CANCELLATION OF YOUR SCORE(S).

PLEASE MAKE SURE THAT YOU HAVE WRITTEN ALL THE REQUIRED INFORMATION ON THE ANSWER SHEET BEFORE YOU WORK ON THE EXAMINATION.

DO NOT TURN THIS PAGE OVER UNTIL YOU ARE TOLD TO DO SO.