Advances in Language and Literary Studies
ISSN: 2203-4714
www.alls.aiac.org.au

Item Analysis of a Multiple-Choice Exam

Sibel Toksöz, Ayşe Ertunç

School of Foreign Languages, Mehmet Akif Ersoy University, İstiklal Yerleşkesi, 15100 Burdur, Turkey
Corresponding Author: Sibel Toksöz, E-mail: sibelbayir@mehmetakif.edu.tr

ARTICLE INFO

Article history
Received: September 14, 2017
Accepted: November 23, 2017
Published: December 30, 2017
Volume: 8 Issue: 6
Advance access: December 2017

Conflicts of interest: None
Funding: None

Key words:
Assessment,
Item Analysis,
Multiple Choice,
Item Facility,
Item Discrimination,
Distractor Efficiency

ABSTRACT

Although foreign language testing has been subject to some changes in line with the different perspectives on learning and language teaching, multiple-choice items have been considerably popular regardless of these perspectives and trends in foreign language teaching. There have been some studies focusing on the efficiency of multiple choice items in different contexts. In Turkish context multiple choice items have been commonly used as standardized stake holder tests as a requirement for undergraduate level for the departments such as English Language Teaching, Western Languages and Literatures and Translation Studies and academic progress of the students in departments. Moreover, multiple choice items have been used noticeably in all levels of language instruction. However, there hasn't been enough item analysis of multiple-choice tests in terms of item discrimination, item facility and distractor efficiency. The present study aims to analyze the multiple choice items aiming to test grammar, vocabulary and reading comprehension and administrated at a state university to preptory class students. In the study, 453 students' responses have been analyzed in terms of item facility, item discrimination and distractor efficiency by using the frequency showing the distribution of the responses of prepatory students. The study results reveal that, most of the items are at the moderate level in terms of item facility. Besides, the results show that 28% of the items have a low item discrimination value. Finally, the frequency results were analyzed in terms of distractor efficiency and it has been found that some distractors in the exam are significantly ineffective and they should be revised.

INTRODUCTION

Woodford (1980) states that as a part of teaching and learning process foreign language testing is with us for a long time. Foreign language tests were utilized to assess learners' knowledge of grammatical rules, their ability to translate literary texts, paragraphs and vocabulary knowledge before the World War II. However, after the World War II, there was a need for communication and armies started to develop their own strategies to teach languages and they were considerably successful. This success led to focus on oral abilities and there was a shift from focus on literacy skills to oral communication skills (Woodford, 1980).

Although there have been shifts in terms of the emphasis on different language skills, foreign language testing were still utilized to assess learners' grammar, vocabulary and reading until 1960s. In other words, because of the effects of behaviorism and contrastive analysis, testing focused on particular language components such as phonological, grammatical, and lexical comparison between two languages in the earlier periods of English teaching (Brown, 2004). Later, institutions such as Modern Language Association designed an exam including four skills namely reading, writing, listening, and speaking.

In addition to the standardized test, the design of the classroom tests has also undergone some changes in line with the latest developments in foreign language teaching. With the effects of communicative language teaching method, different constructs defined as the entity that is being measured (Davidson, Hudson & Lynch, 1985) have been started to be taken into consideration while designing classroom tests in line with the specific language skills (Brown, 2004).

Regardless of the different perspectives and shift in foreign language testing, multiple-choce (MC) tests have been preferred in educational settings in Turkey. High-stake exams administered by Turkish Republic Assessment, Selection and Placement Center (ÖSYM) such YDS or e-YDS are both in multiple-choice format. Hence, multiple-choice tests have gained much more popularity during the recent years because of their gate keeping effects. In addition to these standardized tests, most of the exams at schools or universities such as midterms and finals are in MC tests format because of the huge numbers of students and heavy schedule of teachers.

Although multiple-choice items are commonly used at university level and other levels of language instruction and other subject areas in Turkey, there hasn't been enough evidence about the item analysis of multiple choice tests in the literature. However, "the quality of a test largely depends on the quality of the individual items" (Oluseyi & Olufemi,

Published by Australian International Academic Centre PTY.LTD.

http://dx.doi.org/10.7575/aiac.alls.v.8n.6p.140

2012, p.240), Therefore, this study attempts to fill this gap by answering the following research questions:

(1) What is the difficulty level (item facility) of each item on the multiple-choiced exam administered to preparatory school students?
(2) What is the discrimination index (item discrimination) of each item on the multiple-choiced exam administered to preparatory school students?
(3) What is the distribution of the response patterns (distractor efficiency) for each of the five options of the items on the multiple-choiced exam administered to preparatory school students like?

LITERATURE REVIEW

Multiple Choice Tests

Although there are different types of assessment tools in line with the construct aimed to be assessed, Öztürk (2007) maintains that multiple-choice (MC) items are commonly used in language testing. Multiple-choice items are described as receptive or selective (Brown, 2004). In other words, administration of these items requires the test takers to choose from a set of responses rather than creating a response themselves. As for the basic structure of a multiple choice item, each multiple-choice item has a stem which acts as a stimulus and several alternatives provided to the test takers to be chosen. Of all these alternatives, there is a key which is defined as the most appropriate response to the stem, and the other alternatives are called distractors.

Öztürk (2007) states, multiple-choice items are mostly preferred by teachers thanks to the fact that they are relatively easy to prepare and practical to administer. Multiple-choice items seem to be reliable compared with other types of tests which are negatively affected by subjectivity (Öztürk, 2007). Additionally, Brown (2004) suggests that multiple choice items provide overloaded teachers with the opportunity of easy and consistent process of scoring and grading. It is also maintained that multiple choice items are easy to prepare because there is a computer program especially designed to prepare multiple choice items testing vocabulary (Coniam, 1997). Moreover, MC tests can be graded easily and quickly thanks to Optical Mark Readers.In his study conducted with 57 ESL graduate students, Tsagari (1994) found out that multiple choice-items were significantly easier and less discriminating than free response tasks.

In addition to standardized tests, there were some studies conducted on the efficiency of using multiple choice items in different skills of the language. As for grammar, Adisutrisno (2008) states that grammar instruction and vocabulary teaching are important components of foreign language teaching, and it is a necessity to use multiple-choice items thanks to their advantage of fulfillment of content validity helping the test taker to perform the behavior which is being tested as long as they are prepared by taking the issues into consideration to facilitate the learning process (Mousavi, 2002). As for listening, multiple-choice listening items were found to be easier than free response items even if when the answer was recorded by the test takers in their L1 (Chinese) (Cheng, 2004). Finally, for speaking, multiple choice items don't seem to be used for testing. However, in YDS which is a standardized proficiency test used as a pre-requisite in most of the institutions, there are some questions aiming at testing the test takers' communicative competence by requiring them to choose the most appropriate responses in the given situation. Although these questions are designed to test oral proficiency, they seem to be lack of authenticity defined as "the degree of correspondence of the characteristics of a given language task to the specialties of the task in real life context" by Palmer and Bachman (1996) and content validity.

Although multiple-choice items have been commonly used in foreign language testing, they have some disadvantages. It seems to be relatively easy to design multiple choice items; however, it is extremely difficult to design them correctly (Brown, 2004). Hughes (2003) emphasizes some weaknesses of multiple-choice items stating that this technique only tests recognition knowledge which is a lower mental skill according to Bloom's taxonomy (1956). Moreover, guessing might have a remarkable effect on test scores. On the other hand, multiple choice tests are limited in terms of what can be tested. Another disadvantage of multiple-choice tests might be the challenge to write successful items.

Hughes (2003) introduced another disadvantage of MC tests: "harmful washback". Washback is defined as the concept as consequential perspectives related to score meaning and the intended and unintended consequences of assessment utilization. Alderson and Wall (1993) maintain the fact that influence of washback is generally observed in behavioral and attitudinal changes in teachers and learners that are linked to the introduction of tests bearing important educational consequences. As Qi (2005) claims, high stakes tests produce considerably influential washback in terms of having strong effects on teaching and learning. Hamp&Lyons (1997) also claim that standardized tests hinder the use of the methods for quality education and multiple choice items are regarded as one of the most significant factors making the standardized test have negative washback effect.

Karabulut (2007) conducted a study on the washback effect of the university entrance exam namely, YDS. The study included high school students, teachers and university students. Questionnaires and interviews were used as data collection method and the results show that majority of the students don't find YDS efficient enough to be able to be competent in other language skills. Based on the findings of this study, it is suggested to change the structure of the YDS exam.

Item Facility, Item Discrimination, Distractor Efficienc

There are some concepts to determine the extent each multiple choice item serves to the testing aims. These are item facility, item discrimination and distractor efficiency. The concept of item facility is defined as the extent to which an item is easy or difficult for a determined group of test takers (Brown, 2004). According to Ding and Beichner (2009) it is "a measure of easiness of an item" (p. 2). Bodner (1980) states "these data allow one to determine whether questions

that one feels are trivial are truly trivial, or whether a question is difficult or truly impossible" (p. 189). If an item is too easy or difficult for a group of test takers, it means that it doesn't make a distinction between high ability and low ability group of test takers. This value can be calculated by using the formula stated below.

$$IF = \frac{\text{\# of Ss answering the item correctly}}{\text{Total \# of Ss responding to that item}}$$

An appropriate multiple choice item usually has IFs that range between.15 and.85 (Brown, 2004).

Another concept about the efficiency of multiple choice items is item discrimination. This concept refers to the extent to which an item differentiates between high and low ability test takers (Brown, 2004). Gajjar et al, (2014) define item discrimination as "the ability of an item to differentiate between students of higher and lower abilities" (p.18). If an item gets correct answers from most of high ability group of test takers and incorrect answers from most of the low ability test takers, this means that this item discriminated between the low and high ability group of test takers. This value can be calculated by using the formula below:

$$ID = \frac{\text{high group \# correct} - \text{low group \#correct}}{\text{½ X total of your comparison groups}}$$

Highly discriminating items have a value which is close to perfect 1.0 and the items with low item discrimination have a value closer to zero (Brown, 2004).

The final term distractor efficiency is about how the responses are distributed to the distractors. If a distractor is not chosen as the correct answer by any of the low ability group members, this means it is not an efficient distractor. There isn't a specific formula to find this value, but it is possible to make a conclusion by looking at the frequency table showing the distribution of the responses.According to Malau-Aduli and Zimitat (2012) "a distractor that fails to attract any examinees is dysfunctional, does not assist in the measuring of educational outcomes, adds nothing to the item or the test (psychometrically) and has negative impact upon learners" (p.927).

Besides these item characteristics, Adisutrisno (2008) also states that the two cognitive views should be taken into consideration while designing multiple choice items. These views are Human Information Processing Theory (Carroll, 1986) and Hypothesis Testing Model by Naom Chomksy (1965). According to Naom Chomsky (1965), people are born with language acquisition device which refers to special abstract mechanisms and thanks to these mechanisms children are able to make hypothesis about how the grammar rules work. Based on this theory, Adisutristino (2008) states that, the forms which aren't available in the target language shouldn't be presented in the multiple choice distractors just for the sake of giving four or more distractors, because they might mislead the students. Secondly, Adisutristino (2008) refers to Human Information Processing Theory (Carroll, 1986). The theory suggests that people solve a problem by utilizing the relevant information which has been stored in long term memory. In line with this theory, if a multiple choice item includes two problems, it doesn't facilitate the learning process. Based on this concept, Adisutristino (2008) concludes that a multiple choice item to test grammar should focus on one specific form. Otherwise it seems to be likely to confuse the students rather than testing their proficiency. In addition to grammar, multiple-choice is the most popular method of testing reading comprehension (Shadehah, 1997).

METHOD

This part includes information on participants, instrument, time period and procedure, data analysis and assumptions.

Participants

In this study 453 students studying at a state university in language preparation classes were included. At the beginning of the fall semester, students took an online placement test and they were placed in C level which means A1 in Common European Framework (CEF). In the program students are provided with 26 hours of Main Course and 4 hours of Grammar classes. There are not specific classes for Reading&Writing and Listening&Speaking skills. The writing and speaking parts in the Main Course books are covered in the class. The students do not read the books in the class. However, some activities related to the books are covered in the class.

The exam questions were grouped as A and B. Cluster sampling method was used to choose the participants and the students taking the booklet A were chosen. The questions were the same for each group but the questions were assigned to different numbers and places. The genders of the participants were ignored as the gender was not regarded as an independent variable.

Instrument

Data was collected from the responses given by the participants to 50 multiple choice items constituting a component of the first midterm. The multiple choice part included three main sections: vocabulary, grammar and reading. Vocabulary part was based on the words covered in Main Course class. The grammar part was constituted of the grammar items covered in Grammar classes. The reading passages were chosen according to the students' level. The reading questions were comprehension questions. In addition to the multiple choice part, there were listening, writing and speaking parts. These parts have equal weights in the overall score. To use the results, the necessary permission was taken from the head of the School of Foreign Languages.

Time Period and Procedure

Data was collected in the fall semester in 2014. The exam was administered by the English instructors of the university and the students were supposed to answer the questions in 60 minutes.

Data Analysis

Data was analyzed through the statistics program IBM SPSS Version 20. In order to get results for the dependent vari-

ables frequency analysis was done. To find out the results for item facility (IF), item discrimination (ID), and distractor efficiency; the formulas offered by Brown (2004) were used. The items of the multiple choice part of the exam were analyzed in terms of item discrimination and item facility based on the ranges mentioned in the literature review part (Brown, 2004).

Assumptions

The students were assumed to have the necessary background knowledge to understand the items and answer the questions. They are also assumed to have the same schooling background in terms of English language education. They are assumed to have started to learn English from 4th grade until 12th grade with the curriculum offered by Turkish Ministry of Education. Also it is assumed that, gender is not a significant independent variable in line with the research focuses mentioned in the data analysis part.

RESULTS AND DISCUSSION

In this section the results of the study are presented in three main sub-titles as Item Facility (IF), Item Discrimination (ID), and Distractor Efficiency.

Item Facility

In order to find out the IF for each item the formula mentioned in the literature review (Brown, 2004) was used. According to the results, most of the items have been found to have a moderate level of item facility ranging between.24 and.85. The results show that, 23 of the items (46 %) range between.24 and.60, so they are relatively difficult. And 18 of the items (36 %) range between.60 and.85, so they are relatively easy. So, these items can be claimed to serve to the testing aim.

On the other hand, two of the items have been found to be very difficult. For example, the IF of item # 28 is.11. This means that the item is over the moderate level. Also, the IF of item # 17 is.07 which means that the item is too difficult for the participants. As Brown (2004) suggests, these items can pose a challenge for the high achieving students. Since this number is not very high, they do not seem to have the potential to create negative washback effect on participants.

According to the results some of the items have also been found to be very easy. The Item Facility values of them are as follows: item # 6 =.92, item # 21=.94, item # 23=.92, item # 24=.98, item # 25 =.87, item # 40=.93, and item # 41=.86. As Brown (2004) suggests, these easy items can function as warm-up items and motivate the low achieving students. In these terms, these items might lead the students to have the feeling of success. As a result of this, these items have positive washback effect as Alderson and Wall (1993) suggest.

Item Discrimination

To find out the item discrimination, the formula mentioned in the literature review was used. According to the results, 14 of the items (28%) have been found to have a moderate item discrimination value which is.50 and higher than it. For example item #42 has.67 item discrimination values as shown in the Table 1.1.

The rest 36 items (72 %) have been found to have an item discrimination value which is lower than.50. Table 1.2 shows these items' item discrimination value.

As shown in the table 1.2, some of the items have failed to discriminate the high and low ability students. These items might have negative washback (Hughes, 2003) on high-ability students.

Interestingly, item #17 has been found to have.-09 discrimination value. Table 1.3 shows the distribution of the responses.

As seen in the Table 1.3, high ability students have failed to answer the item correctly contrary to the expectations. That means, this question should be revised in order to prevent negative washback effect.

Table 1.1. Distribution of the responses for Item #42

Item # 42	#Correct	#Incorrect
High-ability Ss (Top 151)	140	11
Low-ability Ss (Bottom 151)	38	113

Table 1.2. Item discrimination values of the items that are lower than 0.50

Item	ID value	Item#	ID value
Item # 1	0.41	Item # 22	13
Item # 2	0.29	Item # 23	11
Item # 3	0.26	Item # 24	05
Item # 4	0.43	Item # 25	19
Item # 5	0.27	Item # 28	17
Item # 6	0.14	Item # 30	29
Item # 9	0.43	Item # 31	33
Item # 10	0.17	Item # 32	35
Item # 11	0.42	Item # 33	42
Item # 13	0.47	Item # 36	49
Item # 14	0.25	Item # 40	09
Item # 15	0.09	Item # 41	21
Item # 16	0.33	Item # 43	33
Item # 17	–0.09	Item # 44	37
Item # 18	0.19	Item # 45	0.45
Item # 19	0.90	Item # 47	33
Item # 20	0.26	Item # 49	39
Item # 21	0.09	Item # 50	32

Table 1.3. Distribution of the responses for Item #17

Item #17	#Correct	#Incorrect
High-ability Ss (Top 151)	4	147
Low-ability Ss (Bottom 151)	19	132

Distractor Efficienc

To analyze the distractor efficiency of the items the frequency results were used. According to the results, Table 2.1 shows how responses across high and low ability students are distributed for item #42.

As seen in the Table 2.1, the distractors of the item # 42 seem to be efficient, because there is an even distribution of the responses from the low ability students. Table 2.2 shows how responses across high and low ability students are distributed for item #17

As seen in the Table 2.2, Distractor A attracts more responses from high-ability students. Significantly few students (4) chose the correct answer (D). However, interestingly more low-ability students chose the correct answer than the high-ability group. This might also lead to negative washback effect on high ability students (Hughes, 2003). The correct responses from the low ability students might be attributed to the chance factor which is seen as a disadvantage of multiple choice items (Hughes, 2003).

According to the results some distractors seem to be less efficient and they have not attracted any participants. For example Table 2.3 shows how responses across high and low ability students are distributed for item #4

As seen in the table, Distractor D is not efficient enough to distract the students. It is significantly important that it has not attracted any responses from high ability students and it has got only one response from low ability students. This item can be omitted or revised in order to distract more students.

Table 2.4 shows how responses across high and low ability students are distributed for item #11

As seen in the table above, Distractor A could not attract any students from any groups. So, the distractor can be revised. Or the item can be edited with only two distractors.

Table 2.5 shows the distribution of responses from high and low ability students for item # 21.

According to the results shown in the table, Distractors A and C do not seem to be efficient enough to distract the students. They should be revised or the item should have only two distractors.

There is another significant item in terms of distractor efficiency. Table 2.6 shows the distribution of responses from high and low ability students for item # 24.

As seen in the table, Distractor A and C have not got any responses from high ability students. Also Distractor B has not got any responses from any groups. So, it can be concluded that the distractor is too easy or irrelevant. That's why, it should be edited. Table 2.7 shows the distribution of responses from high and low ability students for item # 25.

As shown in the table above, the Distractor A seems to fail to attract any responses from high and low ability students. Although other distractors B and C seem to be relatively efficient compared to A, they do not meet the criteria in terms of distractor efficiency considering the number of the participants. All the inefficient distractors are consistent with the statement that all multiple choice items do not necessarily serve to the testing aims in terms of being written just for the sake of providing the students with four or more options (Adisutristino 2008). In line with this, some multiple choice items can have only two distractors.

CONCLUSUION

The purpose of the study was to analyze a multiple-choice exam administred to preparatory school students in tems of item analysis namely item facility, item discrimination and distractor efficiency. Overall analysis of the items, have

Table 2.1. Distribution of the responses for Item # 42

Choices	A	B	C	D
High-ability Ss (Top 151)	2	140	5	4
Low-ability Ss (Bottom 151)	31	38	48	34

*B is the correct response

Table 2.2. Distribution of the responses for Item # 17

Choices	A	B	C	D
High-ability Ss (Top 151)	79	3	65	4
Low-ability Ss (Bottom 151)	70	9	53	19

*D is the correct response

Table 2.3. Distribution of the responses for Item # 4

Choices	A	B	C	D
High-ability Ss (Top 151)	4	127	20	0
Low-ability Ss (Bottom 151)	1	62	87	1

*B is the correct response

Table 2.4. Distribution of the responses for Item # 11

Choices	A	B	C	D
High-ability Ss (Top 151)	0	15	100	36
Low-ability Ss (Bottom 151)	0	104	36	11

*C is the correct response

Table 2.5. Distribution of the responses for Item # 21

Choices	A	B	C	D
High-ability Ss (Top 151)	0	146	0	15
Low-ability Ss (Bottom 151)	2	132	1	16

*B is the correct response

Table 2.6. Distribution of the responses for Item # 24

Choices	A	B	C	D
High-ability Ss (Top 151)	0	0	0	151
Low-ability Ss (Bottom 151)	1	0	7	143

*D is the correct response

Table 2.7. Distribution of the responses for Item # 25

Choices	A	B	C	D
High-ability Ss (Top 151)	0	1	146	5
Low-ability Ss (Bottom 151)	0	15	116	20

*C is the correct response

drawn us to the conclusion that, the multiple choice items seemed to be efficient in terms of item facility. Most of the items had acceptable item facility indexes which mean that the difficulty levels of the items were suitable for the students. On the other hand, although the responses seemed to be distributed evenly, there were some responses which were not discriminating enough between high and low ability students. These items were found to need revision to improve the discriminatory power and the quality of the exam overall. By doing so, the potential negative washback effect of the exam for high ability students could be diminished or even inhibited.The results of the analysis have also identified that there were some distractors which seemed to be completely inefficient. It is suggested that these distractors should be revised or replaced with new alternatives for future use. It is also plausible to suggest that, the teachers or instructors should have some in-service seminars on testing since most of the teachers are supposed to construct some multiple-choice items at some points in their teaching life. It is also suggested that there might be a specific unit responsible for testing and the analysis of the items after the exam; that would be overwhelming for teachers with large enrollments and busy schedules. If the analysis of the items is utilized, positive washback effect can be ensured to the institution as well as the students. For further research other exams such as quizzes could also be analyzed to enable a broader picture. Moreover, the students could be interviewed about the difficulty level of the exam, and discriminatory power of the exam, and the efficiency of the distractors. By doing so, the quantitative results could be strengtened by qualitative data as well.

REFERENCES

Adisutrisno, W. D. (2008). Multiple Choice English Grammar Test Items That Aid English Learning for Students of English as a Foreign Language. *k@ta, 10 (1), 36-52*. doi: 10.9744.

Alderson, J. C. & Wall, D. (1993). Does washback exist?. *Applied Linguistics, 14*, 115-129.

Bachman, L. & Palmer, A. S. (1996). *Language Testing in Practice*. New York: Oxford University Press.

Bloom B.S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay Co Inc.

Bodner, G. M. (1980). Statistical Analysis of Multiple-Choice Exams. *Journal of Chemical Education*, *57*(3), 188-90.

Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. White Plains, NY: Pearson Education.

Carroll, D. W. (1986). *Psychology of language*. Belmont: Wadsworth.

Chomksy, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Coniam, D. 1997. A preliminary inquiry into using corpus word frequency data in the automatic generation of English language cloze tests. *CALICO Journal, 14 (2-4),* 15-33.

Davidson, F., Hudson, T. & Lynch, B. (1985). Language Testing: Operationalization in classroom measurement and L2 research. In Marianne Celce Murcia (Ed.) *Beyond basics: Issues and research in TESOL*. Rowley, MA: Newbury House.

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics-Physics Education Research*, *5*(2), 1-17.

Gajjar, S., Sharma, R., Kumar, P., & Rana, M. (2014). Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. *Indian Journal of Community Medicine*, *39*(1), 17.

Hamp-Lyons, L. (1997). Washback, impact and validity: ethical concerns. *Language Testing,* 14 (3), 295-303.

Hughes, A. (2003). *Testing for language teachers*. Second Edition. Cambridge: Cambridge University Pres.

Karabulut, A. (2007). *Micro level impacts of foreign language test (university entrance examination) in Turkey: A washback study* (Master's Thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1448691)

Malau-Aduli, B. S., & Zimitat, C. (2012). Peer review improves the quality of MCQ examinations. *Assessment & Evaluation in Higher Education*, *37*(8), 919-931.

Mousavi, S. A. (2002). *An encyclopedic dictionary of language testing*. Third edition. Taiwan: Tung Hua Book Company.

Oluseyi, A. E., & Olufemi, A. T. (2011). The Analysis of Multiple Choice Item of the Test of an Introductory Course in Chemistry in a Nigerian University. *International Journal of Learning, 18*(4), 237-246.

Öztürk, M. (2007). Multiple-Choice Test Items of Foreign Language Vocabulary. *Eğitim Fakültesi Dergisi, 20 (2),* 399-426.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing, 22(2),* 142-173.

Shehadeh, M., A. (1997). *The Effect of Test Type on Reading Comprehension in English as a Foreıgn Language: The Case of Recall Protocol and Multiple Choice*. (Doctoral Dissertation). Retrieved from UMI Microform. (9731711).

Woodford, E., P. (1980). Foreign Language Testing. *The Modern Language Journal*, *64 (1),* 97-102.