Investigating the ITU Semi-Computerized MCQ Model: A Systematic Assessment of EFL Students' Attitudes and Testing Behaviours in the Newly Adopted E-Testing Form

Hind BRIGUI

Faculty of Letters and Humanities, Ibn Tofail University, Kenitra, Morocco

Abstract: *The multiple-choice questions exam is, nowadays, widely used especially in large-scale language testing. However, many people criticize this kind of tests as lacking validity and not fully ensuring reliability, as well as because of its negative backwash effect on foreign language teaching and learning. This article is divided into two main sections. Section one takes a look at the concepts of validity and reliability in relation to testing, more exactly as they are approached in an MCQ exam. It also presents some general views on the advantages and disadvantages of the MCQ mode of testing. The second section reports on an empirical research into the attitudes of Moroccan EFL university students towards MCQ tests. The study, which is of an exploratory ex-post facto design, is addressed to explore the learners' testing mode preference, as well as to look at their attitude towards the newly adopted MCQ test. Several issues related to this MCQ form of test are investigated in an attempt to come up with a general multi-faced view of the topic. 150 university students from three proficiency levels represent the subjects of the study. One major set of data yielded from a 22 item questionnaire is examined. Data are quantitatively analyzed through frequency distribution counts. Findings show that MCQ exams are most preferred by students as a testing mode. Yielded results also show an interrelationship between students' proficiency level and their testing mode preference. However, all the groups were found not to differ statistically in their general attitudes towards the fairness, simplicity and acceptability of MCQ tests, as well as their influence on their study habit. The electronic/soft feature of the currently adopted MCQ test shows, surprisingly, an ambivalent attitude of students who think computerized correction can yield a fair grade, but feel more comfortable with a professor's correction.*

Keywords*: Multiple choice question exam (MCQ); Learners' Attitudes; testing behaviours; Computerized correction; Washback effect.*

I.   **Introduction**:

Testing has always been a crucial aspect of education. That is why evaluation is one of the necessary modules of each curriculum development program. Different testing approaches and modes have been developed and adopted for years. This paper is concerned about one specific type of testing which is characterized as a large scale testing facility, the multiple choice questions test (MCQ test, henceforth). This testing mode was first developed during the 1960's influenced by the aural-oral approach and the structuralist linguistic approach against the conventional language testing approaches approaches which were not scientific and introduced psychometric-structuralist linguistic approaches. Lado (1961), as well as his followers, believed that language consists of different parts such as phonetics, vocabulary and syntax, and different skills such as speaking, listening, reading and writing. Each different skill can be broken down into their component parts and those parts tested separately and adequately (Farhady,1979). Such a rationale is derived from structural linguistics and behavioristic psychology which claim that behavior is in essence the sum of its component parts. It is claimed that by adequately sampling the component parts of language, validity can be achieved. It is just under such background that multiple-choice 'items' came into being.

Very often, the multiple-choice item is made up of two parts. The initial part of each multiple-choice item is known as the stem; the choices from which the students select their answers are referred to as options. One option is the correct answer, while the other options are distractors. MCQ exams are now very popular and widely used in national and international examinations such as TOEFL, GRE, GMAT, and have recently been introduced in the Moroccan university, Ibn Tofail University (ITU) as a case. However, as there are many professors, educators and language specialists who are for MCQ exams, the voice of criticism against using this kind of testing is also increasingly growing. This paper is divided into two main sections. The first one is dealing with general issues concerning MCQ exam, wheras the second is a report on an empirical study that targeted ITU EFL students' attitudes about MCQ exams.

1. A review on evaluating MCQ tests
1.1. Validity of MCQ tests

Validity and reliability are the two basic criteria that should be looked at when evaluating any test. They are both interrelated and depending on each other. Validity is the quality which most affects the value of a test, prior to, reliability. The validity of a language test is established by the extent to which it generally succeeds in providing an accurate concrete representation of an abstract concept, for example, proficiency, achievement, or aptitude. (Davies, A. et al, 2002 : 221). Validity was defined as It is traditionally defined as "the degree to which a test measures what it claims, or purports, to be measuring" (Brown, 1996, p. 231), or the extent to which scores obtained on an assessment instrument represent true knowledge (Cook & Beckman, 2006). Validity consists of different types. Face validity, predictive validity, concurrent validity, construct validity and content validity.

Face validity means whether the test looks like what it is supposed to test. So it is also called surface credibility or public acceptability. Direct tests have higher face validity than indirect tests. Predicative validity deals with whether the test can predict the students' performance in the future, for example in a placement test. Concurrent validity means comparison between a test and an established test. It is concerned of how well a new test compares to a well-established test. It can also refer to the practice of concurrently testing two groups at the same time, or asking two different groups of people to take the same test (Bachman, 1997).

Construct validity refers to the experimental demonstration that a test is measuring the construct it claims to be measuring. A construct, or psychological construct as it is also called, is an attribute, proficiency, language ability, competence or skill that happens in the human brain and is defined by established theories. Content validity is about how much the test can measure the knowledge and abilities the test takers should attain. It refers to representativeness of the content selected for assessment. In accordance with the definitions and theories of validity above, it was suggested that MCQ tests are of very low validity because they lack face validity, concurrent, construct and even sometimes content validity. (Bachman, 1997)

1.2. Reliability of MCQ tests

Test reliability refers to the consistency of scores students would receive on alternate forms of the same test. Due to differences in the exact content being assessed on the alternate forms, environmental variables such as fatigue or lighting, or student error in responding, no two tests will consistently produce identical results. This is true regardless of how similar the two tests are. The more comparable the scores are, the more reliable the test scores are (Wells and Wollack, 2003).

It is important to be concerned with a test's reliability for two reasons. First, reliability provides a measure of the extent to which an examinee's score reflects random measurement error. Measurement errors are caused by one of three factors: (a) examinee-specific factors such as motivation, concentration, fatigue, boredom, momentary lapses of memory, carelessness in marking answers, and luck in guessing, (b) test-specific factors such as the specific set of questions selected for a test, ambiguous or tricky items, and poor directions, and (c) scoring-specific factors such as nonuniform scoring guidelines, carelessness, and counting or computational errors. These errors are random in that their effect on a student's test score is unpredictable – sometimes they help students answer items correctly while other times they cause students to answer incorrectly. In an unreliable test, students' scores consist largely of measurement error (Wells and Wollack, 2003).

It is desirable to use MCQ tests with good measures of reliability, so as to ensure that the test scores reflect more than just random error. However, formally assessing the validity of a specific use of a test can be a laborious and time-consuming process. However, there are primarily two factors at an instructor's disposal for improving reliability: increasing test length and improving item quality.

Yet, before lengthening a test, it is important to consider practical constraints such as time limit and examinee fatigue. As a general guideline, it is wise to use as many items as possible while still allowing most students to finish the exam within a specified time limit. On the other hand, the MCQ Item quality has a large impact on reliability in that poor items tend to reduce reliability while good items tend to increase reliability. Items that discriminate between students with different degrees of mastery based on the course content are desirable and will improve reliability. An item is considered to be discriminating if the "better" students tend to answer the item correctly while the "poorer" students tend to respond incorrectly (Wells and Wollack, 2003).

II. Advantages And Disadvantages Of MCQ Tests

MCQ tests are still raising a controversy in defining whether this type of tests is good or inappropriate. Many language professionals and test makers stated different advantages, as well as disadvantages to MCQ exams. Generally speaking, the advantages and disadvantages are mainly as follows

2.1. Advantages of MCQ tests

There is a series of advantages with multiple-choice tests. First, the MCQ exam is is of a high degree of objectivity and fairness and high reliability. It has the strengths of saving time, effort and money in carrying out large-scale examinations. It is also of clear subject and simple answer. it is easy for scorers to go over examination papers. Since it usually has only one correct answer, or at least, a limited number of correct answers, so it can be scored mechanically, and therefore, the reliability of scorers can be ensuredIt is easy to predict. The difficulty coefficient of each item or the whole test can be estimated in advance. Then it can be modified in time to make sure a perfect testing (Bachman and Palmer, 1996). Moreover, it is quite convenient for statistic analysis and can also provide more feedback for language teaching and learning, which makes this kind of test more scientific. In addition to this, the range of the multiple-choice item is wide enough. It can be employed in any disciplines, which strives for the process of dealing with the problems, such as mathematics, chemistry and physics, in English teaching and learning, whether practical courses or theoretical courses such as linguistics, literature, this kind of item is widely used in tests (Luo and Zhang, 2011).

2.2. Disadvantages of MCQ tests

There are many disadvantages with MCQ test, firstly, frequently it does not lend itself to the testing of language as communication. it is also seen as an indirect testing that can't test the real language ability of the candidates. Weir said : "The MCQ is not a valid testing item, because in real life, people rarely use a four-selection way to express if understand or not, and we show our understanding of listening and reading by speaking and writing（Weir, 1990 : 44). Moreover, MCQ test encourages guessing, a factor of guessing is one that can"t be excluded in the process of choosing the answer in a test paper. In fact, in a question with four answer options, the chance of a student guessing an answer is 25%, even if candidates are ignorant of the four options, and also the chance will surely be higher if they can eliminate one or two of the distractors. Another disadvantage is the complexity of items construction. It is common knowledge that the process of constructing subjective items is easier than that of constructing multiple-choice item. On the contrary, the quality of multiple-choice item is hard to ensure because of this complex process of item writing. At first sight, the construction of multiple-choice item is quite simple, but in fact, it is rather difficult to control its item and alternatives. A number of questions should be taken into consideration, for example, which linguistic elements should be tested, how to design item and key in each item, how to design plausible distractors, and so on, all of which really cost much time and effort (Luo and Zhang, 2011). Finally, it is easier for candidates to cheat in multiple-choice item. The reason for this phenomenon is that candidates are only required to choose the key from four alternatives, so it is quite convenient for some candidates to copy other candidates" answers. Even though from the test papers the scorers can make sure that some candidates have cheated in the exams, however, it is not easy to collect strong evidence in order to punish them (Tsagari, 2007).

III. Studies On Students' Preferences And Attitudes Towards Tests

The existing literature indicates that, in general, multiple-choice and short open-ended questions are most preferred and student scores were most determinant of student preferences regarding tests (Yunker, 1999). It was found that students hold strong views about different formats and methods of assessment. For example, within conventional assessment, multiple choice format exams are seen as favourable assessment methods in comparison to essay type questions. However, Students with good learning skills, who have high confidence in their academic ability, tend to prefer the constructed response type of assessment over the multiple choice type (Zeidner, 1987). Hembree (1988) found that students with high test anxiety, have more unfavourable attitudes towards the MCQ format and a preference to the choice- response type, probably because it puts less demanding requirements on their information processing capacity during the testing situation where that capacity is occupied by worries and test- irrelevant thoughts (e.g. Hembree, 1988). According to him, students report more positive attitudes towards MCQ tests on grounds that these tests are easier to prepare, are easier to take, and hold forth hope for higher relative scores. Traub and MacRury (1990) found similar results. As for as Boud and Falchikov (2007), student's likes and dislikes for particular test modes have changed very little in recent years. Multiple-choice items were in the past and still are generally recognized by students as the most widely applicable and useful mode of evaluation. Similarly, they maintain that testing practices and ultimate decision-making processes among professors have not changed extensively. This study was designed to explore the Moroccan EFL learners's testing mode preference, as well as to look at their attitude towards the newly adopted MCQ test. Several issues related to this MCQ form of test are investigated in an attempt to come up with a general multi-faced view of the topic. To achieve this aim, six research questions have been formulated for the investigation to be carried out.

Research questions

1- Do Moroccan EFL ITU students prefer MCQ exams over the other types of tests?

2- Do ITU students think MCQ exams really reflect their degree of competence?

3- Do ITU students think that their answers on MCQ tests are more affected by their test performance approach than their knowledge of the subject?

4- Is the soft/electronic facet of MCQ test in ITU a comforting factor for students?

5- Is there a relationship between the students' level of proficiency/class level and their testing mode preference?

6- Does the introduction of MCQ tests into the Moroccan ITU context have an impact on the learning habits of students?

IV. Study Methodology

4.1. Subjects

Subjects of this study are 150 students studying English at the faculty of letters and humanities in Ibn Tofail University in Kenitra, Morocco. The variables of age or gender were not important in this study; however, proficiency level, also termed class level in the study, was an important variable in the study. Partcicipants belong to different class levels namely; Semester 2 (N. 36), semester 4 (N. 63) and semester 6 (N. 51). The reason for this choice is to try to get a more representative view of different students' attitudes and opininons towards the investigated topic.

Number of Students	Number (N)	Percentage (%)
Semester 2	36	24
Semester 4	63	42
Semester 6	51	34
Total	150	100

Table 1: The overall number and class level of respondents
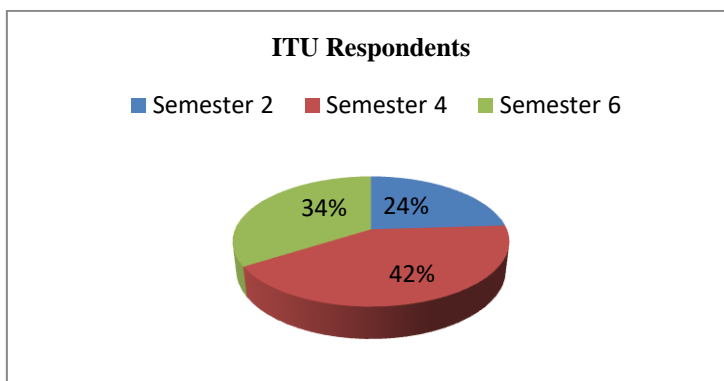
Graph 1: ITU Respondents

4.2. Research instrument

As a research instrument, a questionnaire was designed in a way to elicit relevant information about two main issues; the students' testing mode preference, and their attitudes towards MCQ exams in particular. The questionnaire comprises 22 questions, each set is intended to answer a given research question.

V. Results And Discussion

The research has a quantitative design since the questionnaire was analyzed and made to yield numerical data that were conducted to statistics. Findings are presented as follows to answer each research question.

1- Do Moroccan EFL ITU students prefer MCQ exams over the other types of tests?

Students' reports on their exam type preference are presented in table 1. The data indicate that MCQ tests were the most preferred mode of evaluation with a frequency of 56,6 %, whereas the two other types of tests were lower in the preference frequently.

	Essay/Paragraph questions	Questions about definitions of concepts	Multiple-choice questions	Total Number (N)
Number of preferences	34	31	85	150
Percentage %	22,7	20,7	56,6	100%

Table 2: Students' test type preferences

As for the reasons that students gave for preferring MCQ exams, 77 (51,3%) of the participants agreed that they are easy to prepare for, while 64 (42,7%) agreed that they are easy to take.

2- Do ITU students think MCQ exams really reflect their degree of competence?

	Agree		Neither agree nor disagree		Disagree	
	N	%	N	%	N	%
MCQ exams are fair	69	46	34	22,7	47	31,3
MCQ Exams yield high grades	67	44,7	37	24,6	46	30,7

Table 3: Students' attitudes about MCQ exams and their degree of competence

69 students which is 46% of the total number believe that MCQ exams are really fair in reflecting their degree of competence. They also think that, with MCQ exams, they better stand a chance to get higher grades (N= 67/44,7%).

3- Do ITU students think that their answers on MCQ tests are more affected by their test performance approach than their knowledge of the subject?
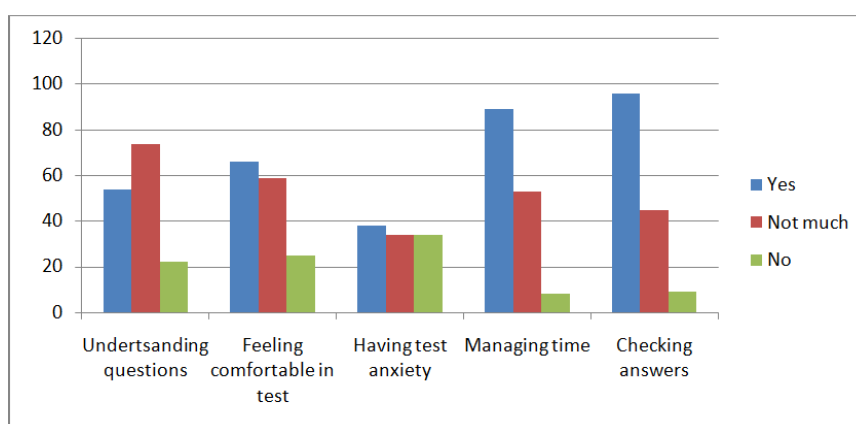
Graph 2: Approaching test performance issues

This question addresses different aspects that students go through while taking MCQ exams, namely, understanding the question's items (71, 45,4% feel comfortable), feeling comfortable in the exam (66, 42,5%), experiencing test anxiety during the exam (38, 35,8%), managing the allotted time (89, 55,7% always have enough time), and checking the answers at the end of the test (96, 59,8%). Approaching the MCQ test, students feel more concerned with their test performance issues rather than with their knowledge of the subject matter. Here, respondents, for example, stress on characteristics that bother them more like their low understanding of the MCQ items, their discomfort taking an MCQ test, and their experiencing test anxiety. However, a great number of students state that they have no problem with the limited allotted time which they find sufficient enough to even check their answers at the end of test.

4- Is the soft/electronic facet of MCQ test in ITU a comforting factor for students?

	Comfortable		Average comfortable		Uncomfortable	
	N	%	N	%	N	%
Being corrected by a computer software	46	30	52	35	52	35

Table 4: Attitudes of students on the software correction

	Yes		Neutral		No	
	N	%	N	%	N	%
Electronic correction can yield a fair grade	59	40,6	42	26,8	49	32,9

Table 5: Attitudes on students on the fair grading

Addressing the soft/electronic facet of MCQ test, results show a quite contradicting finding. While 59 students (40,6%) believe that electronic correction can yield a fair grade, 52 students (or 35%) are uncomfortable being corrected by a computer software. Reasons for this, as discovered in the researcher's personal interactions with respondents, are that students prefer the human contact of their professor for a more 'subjective' 'forgiving' and 'tolerant' correction.

5- Is there a relationship between the students' level of proficiency/class level and their testing mode preference?

	Essay/Paragraph questions		Questions about definitions of concepts		Multiple-choice questions	
	N	%	N	%	N	%
Semester 2	2	5,5	4	11,2	30	83,3
Semester 4	11	17,5	9	14,3	43	68,2
Semester 6	21	41,2	18	35,3	12	23,5

Table 6: The relationship between testing mode preferences and the students' proficiency level

Results show that there exists a relationship between the students' proficiency/class level and their testing mode preference. Semester 2 students show a total preference of 83,3% for MCQ exams, while only 5,5% of their total number go for essay/paragraph questions exam type. Semester 4 students similarly display a preference of 68,2% for MCQ tests, whereas semester 6 students' total preference didn't exceed 23,5%, preferring the essay/paragraph question type with a percentage of 41,2. This shows that as learners get more proficient in the language, they prefer more subjective, open, complex and elaborate testing modes.

6- Does the introduction of MCQ tests into the Moroccan ITU context have an impact on the learning habits of students?

	Agree		Neither agree nor disagree		Disagree	
	N	%	N	%	N	%
MCQ exams influence study method	88	58,7	37	24,7	25	16,7
Students study based on previous final exams	83	55,3	35	23,3	32	21,3
Students practice MCQ solving old exam samples	93	62	33	22	24	16
Students spend less time studying sections unlikely to be an MCQ exam	62	41,3	43	28,7	45	30
Students study all important points without considering final MCQ exam	55	36,7	23	15,3	72	48
Students review previous exams with professors	47	31,3	35	23,3	68	45,3
Final exams provide good feedback for students' studying	83	55,3	28	18,7	37	24,7

Table 7: The impact of MCQ on students' learning habits

The last finding shows that the adoption of MCQ test into the Moroccan ITU context has an impact on the learning habits of students. Respondents show it clear that MCQ exams have influenced their study method (N= 88/58,7%). They also agree that new habits have been formed related to MCQ exam mode, namely, practicing MCQ solving old exam samples (N= 93/62%), focusing on lesson sections that are likely to be in an MCQ exam (N= 62/ 41,3%), or reviewing previous MCQ exams with professors in class (N= 47/ 31,3%). This finding is a clear evidence that the new application of MCQ exam in large scale examinations produces a washback effect on language learning and teaching, which should be taken into account by professors in designing syllabi that present new methods, and enhance ways in dealing with multiple choice questions.

VI. Conclusion

As one of the most popular language testing modes, MCQ exam deserves to be explored not only from theory but also from practice. The fact we should bear in mind is that the MCQ has more disadvantages over advantages. Its big advantage is the objectivity of scoring and economy in terms of manpower, time and material resources. But the complex process of constructing items, especially the writing of distracters, is much more difficult than expected. For large-scale contexts, MCQ is to be used not because it is the best one, but because there are no other choices at present. Therefore, in order to improve the validity of multiple-choice items, the tests in large-scale environments should be made up of two parts, multiple-choice questions on one hand and a few of subjective questions (for example, composition, translation, paraphrasing) on the other, with the aim of making full use of the strengths of both subjective and objective questions to guarantee the quality of tests. Thus such a set of questions will have both good reliability and validity, and also will produce good backwash effect on teaching.

Bibliography

[1]   Bachman, L. F. (1997). Fundamental Consideration in Language Testing. Oxford: Oxford University Press.
[2]   Bachman, L and A. Palmer, (1996). Language Testing in Practice. Oxford: Oxford University Press.
[3]   Boud, D., & Falchikov, N. (Eds.). (2007). Rethinking Assessment for Higher Education: Learning for the Longer Term. London: Routledge.
[4]   Brown, J. D. (1997). 'Testing washback in language education'. *PASAA Journal, 27*, 64-79.
[5]   Cook, D. A., & Beckman, T. J. (2006). 'Current concepts in validity and reliability for psychometric instruments: Theory and application'. *The American Journal of Medicine, 119* (2), 166.e7-166.16. doi:S0002-9343(05)01037-5 [pii]

[6]   Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (2002). Dictionary of Language Testing. Cambridge: CUP/UCLES.
[7]   Farhady, H. (1979). 'The Disjunctive Fallacy between Discrete-Point and Integrative Tests'. In *TESOL Quarterly,* Vol. 13, No. 3, pp. 347-357.
[8]   Hembree, R. (1988). 'Correlates, causes, effects, and treatment of test anxiety'. *Review of Educational Research, 58,* 47-77.
[9]   Lado, R., (1961). Language Testing: The Construction and Use of Foreign Language Tests. Longman, London.
[10]   Luo, S. and X. Zhang, (2011). 'Multiple-Choice Item and its Backwash Effect on Language Teaching in China', *Theory and Practice in Language Studies*, Vol. 1, No. 4, pp. 423-425.
[11]   Traub, R. E., & MacRury, K. (1990). 'Multiple choice vs. free response in the testing of scholastic achievement'. In K. Ingenkamp & R. S. Jager (Eds.), T*est und Tends 8: jahrbuch der pädagogischen diagnostik* (pp. 128-159). Weinheim und Base: Beltz Verlag.
[12]   Tsagari, D. (2007). Review of Washback in Language Testing: How has been done? What more needs doing? Unpublished thesis: Lancaster University, UK.
[13]   Weir, C. J. (1990). Communicative Language Testing. New York: Prentice Hall.
[14]   Wells, C. and J. Wollack, (2003). An Instructor's Guide to Understanding Test Reliability, Testing & Evaluation Services. University of Wisconsin: Madison
[15]   Yunker, D. B. (1999). 'Adding authenticity to traditional multiple-choice test format'. *Education*, 120, 82-87.
[16]   Zeidner, M. (1987). 'Essay versus multiple-choice type classroom exams: the student's perspective'. *Journal of Educational Research, 80* (6), 352-358.