ESSAY OR FORCED-CHOICE TESTS?
FACTORS DETERMINING INSTRUCTORS' PREFERENCES IN DESIGNING TESTS FOR CORE CURRICULUM REQUIREMENTS IN THE COLLEGE OF BASIC EDUCATION IN KUWAIT

[*1]Ali A. Taqi, [2]Hanan A. Taqi, [2]Khaled M. Shuqair

[*]This research was funded by the Public Authority for Applied Education and Training.

[1]Department of Educational Foundations and Administration, College of Basic Education, Ardhyia, Kuwait. Email: prof.a.taqi@gmail.com

[2]The English Department, College of Basic Education, Ardhyia, Kuwait

ABSTRACT: *This research aims at exploring the factors determining instructors' preferences in designing tests for Core Curriculum required courses at the College of Basic Education in Kuwait. For this purpose, a sample of 120 instructors was randomly chosen from a population consisting of 500 instructors teaching the Core Curriculum Program in the CBE. The sample of instructors responded to a 21-statement questionnaire. Furthermore, a sample of 12 instructors, selected randomly from the population, was interviewed. The results of the study were based on the instructors' response to a questionnaire and the interview related to two Core Curriculum required courses (Science Education and Kuwait and Development) which fairly represent the Core Curriculum Program at CBE. The study has shown that the main factors determining the choice of one test over the other are external to the processes and steps used in test design, like the class size and the amount of time consumed in designing and scoring the tests. Such factors were prioritized by instructors mainly because of special circumstances related to CBE's admission policies of accepting a large number of students without strategically planning for proper teacher-student ratio. The study also showed that the tests were limited to the lower-order thinking skills, knowledge, comprehension and retrieval of information, a shortcoming attributed to the course designers who limited the learning objectives to the lower levels of thinking.*

KEYWORDS: *Testing, education, curriculum, forced-choice tests, essay tests, higher education*

BACKGROUND

This paper aims at examining the factors determining the instructors' preferences in designing tests for the Core Curriculum required courses in the College of Basic Education (CBE) in Kuwait. In its capacity as a teacher-preparation state college, CBE houses eleven degree-awarding departments (in addition to a number of service units), all of which are geared towards graduating students with a B.A. degree in education. In addition to the departmental requirements and electives, the CBE offers Core Curriculum courses which all students joining the college should complete towards earning the B.A. degree. The core curriculum lays the foundation of general education in the natural sciences, social sciences and humanities. The Program carries the total of 24 hours of credit, 18 (6 three-credit-hour courses) of which are Core Curriculum requirements, and 6 (2 three-credit-hour course) Core Curriculum electives.

The Core Curriculum Requirements consist of the following courses:

1. The Arabic language   3 credit hours.

2.Science Education   3 credit hours

3. Islamic Education   3 credit hours

4.Kuwait and Development   3 credit hours

5.English Program   (Two 3-credit hour courses).

The core curriculum requirements and electives are offered and taught by most of the departments in CBE. Since the number of students enrolling in CBE is huge (the 2014-2015 enrollment is 4500 students), the CBE administration tries to cater for the needs of this large number of students by offering more sections teaching the Core Curriculum courses. Still every semester the problem persistently arises that a good number of students cannot take these courses because the sections reach the maximum number. In an attempt to deal with the problem, the CBE administration has taken a step further, which is enlarging the class size in every section from 50 students to 70-80 students per section in the recent years. Of course, this step has imposed more pressure on teachers, the teaching process and the learning experience. On the way, compromises have been made at the expense of the quality of education and academic accreditation, which the College has been seriously and diligently seeking in the last few years. One major compromise that has been made by the instructors of core curriculum courses (especially of the required courses which are highly demanded) is related to the process of designing tests for these courses and types of tests (forced-choice or essay tests) frequently used by instructors to measure the mastery of the learning objectives. This paper attempts to examine the factors determining the instructors' preferences in designing tests for the Core curriculum courses and the rationale underlying such preferences. In other words, the study is interested in the extent to which instructors are aware of the principles governing the choice of type of test format over the other, and therefore, whether their tests are able to measure the learning objectives set for these courses. Furthermore, the compatibility of the test designed by the instructors with the learning objectives of these courses and the ability of the learning objectives to measure the various levels in the cognitive domain fall within the scope of the study.

Of the six Core Curriculum requirements, we chose two courses to be the basis of this study: *Science Education* and *Kuwait and Development*. The results of the study are based on the response of the instructors of these two courses to a questionnaire and an interview. We believe that these two courses are fairly representative of Core Curriculum requirements in the CBE for two reasons. First, the two courses fall within two major domains of knowledge in general education: the domain of the natural sciences and that of social studies and humanities. Consequently, results obtained from responses related to these two courses can be informative of the rest of the Core Curriculum courses. Second, the two courses involve several majors and departments in CBE, which offer and teach them. This means that the sample of the study; i.e., the instructors of the two courses come from several departments, and hence they can be fairly representative of the instructors involved in teaching the Core Curriculum requirements and electives.

Significance of the Study

In an extended review of the influence of testing on the curriculum, Madaus (1988) posits a number of principles. Principle 1, which is relevant to our study, states:

> The power of tests and exams to affect individuals, institutions, curriculum or instruction is a perceptual phenomenon. If students, teachers or administrators believe that the results of an examination are important, it matters very little whether this is really true or false. The effect is produced by what individuals perceive to be the case (p. 88).

Accordingly, what matters is whether participants believe a test is important and it meets the needs set for it. This explains the symbolic power in the minds of policy makers and instructors. In other words, the value of tests lies in how we perceive them and the importance we attach to them, irrespective of the educational process itself (Gipps, 2012, p. 29). The significance of this study lies in exploring the CBE instructors' perceptions about the types of tests they design for the Core Curriculum required courses (forced-choice and/or essay questions), and whether these tests are able to measure the students' mastery of the learning objectives and address various levels of thinking. No doubt, test designs and preference of one type over the other have a lot of implications with regard to the quality education and the competencies that students should master within the main learning objectives of the course. Hence, in order to have a program that meets the international standards of quality education and academic accreditation, this program should also be comprised of methods of assessments that go beyond the lower-order thinking skills and address the higher-order thinking skills. The researchers hope that the study might yield results that would be beneficial to the decision makers and educators in the College of Basic Education (which is currently seeking academic accreditation) and to educators in Kuwait and the Arab world at large.

Research Questions

The study poses four interchangeable questions:

1. What types of tests (forced-choice or essay) do teachers of Core Curriculum courses in the College of Basic Education prefer to design?
2. Are these tests compatible with the learning objectives set for the course?
3. What are the factors determining the instructors' choice of one type of test over the other?
4. To what extent do the tests designed by the instructors of CBE address various levels of thinking on Bloom's taxonomy?

THEORETICAL BACKGROUND

Forced-choice and Essay Tests

In *Transforming Classroom Grades*, Robert Marzano lists seven forms of classroom assessment that can be used to keep track of student achievement on specific topics: forced-choice, essay, short written response, oral reports, performance task, teacher observation and student self-assessment. Of these seven forms, the most common methods of testing are the essay and forced-choice types. Marzano further uses a scale of high (H), Medium (M) and Low (L) to indicate the extent to which a particular form of assessment is well suited to a particular skill or process of the curriculum (pp. 86-7). The scale for forced-choice and essay forms when it comes to aspects of grading and curriculum content is shown in Figure 1 below:

Figure 1: Curriculum content and the scale for forced-choice and essay questions

Skills and Processes Assessed	Assessment	
	Forced-Choice	Essay
Informational Topics	M	H
Process Topics	L	M
Thinking and Reasoning	M	H
Communication	L	H
Non-achievement Factors	L	L

According to Figure 1, essay forms of questions receive the score of H for informational topics, thinking and reasoning and communication. However, the highest score Forced-choice forms receive is M for informational topics and thinking and reasoning. This means that forced-choice forms can be used to asses these skills and processes, but better assessments (basically essay forms) may be used (p. 87).

Indeed, forced-choice items, though more commonly used in standardized tests, are still relied upon in classroom tests to measure the mastery of a wide range of objectives. Forced-choice items can be defined as follows:

> *The classic objectively scored paper-and-pencil test. The respondent is asked a series of questions, each of which is accompanied by a range of alternative response. The respondent's task is to select either the correct or the best answer from among the options. The index of achievement is the number or proportion of questions answered correctly (Stiggins, 1994, p. 84).*

Measurement experts, such as Thomas Haladyna (1994), describe a wide variety of forced-choice items that include: conventional multiple-choice, matching, alternate-choice, true-false, multiple-response and fill-in-the-blank items. Generally speaking, forced-choice items are fairly difficult and time consuming to design. The most difficult aspect of writing forced-choice items is designing viable distracters, which must be inaccurate enough to be considered wrong by students who understand the content of the course, but reasonable enough to be chosen by students who rely on an "educaed guess" (Marzano, p. 88).

As we have seen, essay questions are highly effective tools for assessing informational topics, thinking and reasoning, and communication. However, though essay questions are more likely to measure the thought processes and communication skills (especially the writing skills) they are subject to bias in grading and hence, the responses to these questions are less stable than the forced-choice items. In contrast, forced-choice items are objectively scored and less biased. However, as Carr (2015, p.29) rightly points out that it proves difficult to write forced-choice items without training and experience (p. 29). The difficulty stems from the fact that untrained and inexperienced teachers are most of the time unable to come up with viable distracters.

Whether the testing method used is forced-choice or essay questions, the general tendency in course examinations is to pose the question, 'How much do you remember of what has been covered?' rather than, 'What can you do with what you have learned?" (Marzano 2000, p. 87) While both questions are important, tests should reflect the learning objectives the instructor has set for the course. As Wortham (1990) points out, teacher-designed classroom tests, though less rigorously constructed than standardized tests, must accurately measure objectives for classroom instruction. In other words, tests should be carefully designed to fit the learning objectives (p. 143). Do the objectives require students to recall definitions and recognize facts, to solve problems, or to do both? Do the objectives require students to separate ideas into component parts, to combine ideas into a new product, to judge ideas with established standards, or all three? The course objectives should determine the kind of questions the teacher uses. The questions relating to course objectives involve different levels of thinking and learning (Cameron 1997).

Before a test can be organized to measure the curriculum objectives, it is necessary for the teacher to understand more accurately what skills are to be measured and to what extent the student will be expected to demonstrate mastery of the objective. That is, the test items should reflect the level of understanding that is required to master the objective. Analysis of the level of understanding is commonly done by constructing a *table of specification*. In this table objectives are charted using Bloom's Taxonomy of Educational objectives. The table describes levels of understanding in the cognitive domain ranging from the ability to recall information to the highest level of understanding, which is evaluation. These levels can be described in the following way (Bloom, 1956):

Figure 2: Explanation of Bloom's Taxonomy (Source: Bloom, 1956)

	Level of Understanding	Descriptive Terms Used in Stating objectives	
1	**Knowledge** Simple recognition or recall of material or facts The ability to remember or recognize information	Tell List Name Classify	Define Identify Locate
2	**Comprehension** The ability to translate information in your own words The ability to show that one understands	Restate Discuss Summarize Explain Review	Describe Interpret
3	**Application** Problem solving or applying ideas to new situations The ability to use information or apply learning to new situations and real life circumstances	Demonstrate Dramatize Construct Imply Illustrate	 Practice
4	**Analysis** Separating ideas into component parts and examining relationships Indentifying parts of information and its relationship to the whole	Organize differentiate Compare Distinguish	 relate
5	**Synthesis** Combining ideas into a statement new to the learner The ability to take information from various sources and present it in a created form	Design Compose	Create Develop
6	**Evaluation** Making judgments by using self-produced criteria or established standards. The ability to evaluate based on standards	Decide Conclude	Judge Assess

These six levels can be further divided into lower- and higher-order thinking skills. Lower-order skills involve knowledge and comprehension skills, while higher-order skills involve application, analysis, synthesis, and evaluation skills. The difference between lower and higher order skills is that higher order skills require the active use of course material while lower-order skills do not. Knowing the thinking and learning expectations the course objectives set for students helps the teacher determine the appropriate type of exam questions and provides a guide for how questions should be presented. (Grondlund, 1995). According to the discussion above, in principle both forced-choice and essay items may be used to test a wide range of learning objectives. However, essay types tend to be better tools for assessment to test informational topics and recall and comprehension skills (lower-order thinking skills) and thinking and reasoning skills (higher-order thinking skills). Furthermore, essay forms are a better tool for assessing the written communication skills as they ask students to construct their thought processes in written forms. No doubt, test designs and preference for one type over the other have a lot of implication with regard to the quality of education and the

competencies that students should master within the main learning objectives of the course. Hence, in order to have a program that meets the international standards of quality education and academic accreditation, this program should also be comprised of methods of assessments that go beyond the lower-order thinking skills and address the higher-order thinking skills. Nowadays, this attention paid to higher-order thinking skills has become a main pedagogical goal in Europe and the United States (Marzano 2000, Mindes et.al 1997 and Wortham 1990). Likewise, in the Arab world, educators are calling for this shift from lower-order thinking skills to higher-order thinking skills to be effected with regard to methods of teachings and assessment (see, for instance Abu 'Allam 1987, Muqadam 2002, 'Odah 1993, Jaber 1997 and 'Obaidat 1988). However, the reality of education in the Arab world still shows that recall of information and knowledge are the main focus and teachers tend to shy away from higher levels of thinking skills.

Core Curriculum Requirements in CBE

In this section, *Science Education* and *Kuwait and Development*, the two Core Curriculum courses chosen for this study as a fair representation of the Core Curriculum Requirements offered by CBE, are analyzed in terms of their learning objectives, the levels of thinking and the skills assessed, and hence, the best method of testing that can be used. In undertaking this task, the researchers rely on the theoretical background furnished in the previous section.

Science Education

This course is offered by the Department of Science in CBE. It emphasizes the unifying aspects of the scientific approach to the study of nature and human behavior. Part of the course is devoted to a discussion of the nature of scientific inquiry and investigation. The course also focuses on fact identification and concept formation and testing. The course sets some of the major concepts and theories of science into a broader, philosophical and cultural context and traces the development of these theories and concepts to their present status.

Figure 3: Specifications for the learning objectives of *Science Education*

	Learning objective	Level of Understanding	Skill or process assessed	Type of Test	
				Forced-Choice	Essay
1	Explain the components and dimensions of scientific approach	Comprehension	Informational Topic	M	H
2	Discuss the nature of scientific inquiry and the interrelationships between science and technology, society and the environment.	Comprehension	Informational Topic	M	H
3	Explain basics of the scientific processes used in the scientific approach.	Comprehension	Informational Topics	M	H
4	Explain the basic psychological principles and the different theories of teaching science in schools.	Comprehension	Informational Topics	M	H
5	Classify general and specific objectives of teaching science in accordance with various specializations.	Knowledge	Informational Topics	M	H

4.1.1 Kuwait Society and Development

The course offers a description and analysis of social and cultural characteristics and problems of contemporary Kuwaiti society, taking into consideration the specific historical, economic and ideological forces that shape it. Social basis for Kuwaiti identity is examined and students receive an introduction to basic concepts and principles for understanding social phenomena and development.

Figure 4. Specifications for learning objectives of *Kuwait and Development*

	Learning objective	Level of Understanding	Skill or process assessed	Type of Test	
				Forced-Choice	Essay
1	Explain the various aspects of the meaning of development at the national and global levels.	Comprehension	Informational Topic	M	H
2	Analyze the social and cultural background of Kuwaitis society and the forces that shape it.	Analysis	Thinking and Reasoning	M	H
3	Identify the Kuwaiti geographical boundaries, and the relationship between the people and the land.	Knowledge	Informational Topic	M	H
4	Describe and analysis the effects of the Iraqi invasion on the Kuwaiti society and environment focusing on ways of recovery.	Comprehension	Informational Topics	M	H
5	Describe the main characteristics of population growth of the Kuwaiti society before and after the discovery of oil.	Comprehension	Informational Topics	M	H
6	Identify the features of the Kuwaiti economy before and after the discovery of oil and the impact of the Iraqi invasion	Knowledge	Informational Topic	M	H
7	Describe education development in Kuwait and its impact on the society and the impact of the Iraqi invasion on the Kuwaiti educational institutions	Comprehension	Informational Topic	M	H

   4.1.2   Analysis of the tables of specifications

A reading into the tables of specification for both courses reveals a number of implications. First, the levels of thinking skills the objectives address fall with the domain of lower-thinking skills. That is, the learning objectives of these courses fall short of the higher-order thinking skills, like application, analysis, synthesis, and evaluation. This means that the courses, most probably like the other courses in the Core Curriculum Program of CBE, don't go beyond the recall and comprehension of information. Second, in principle, for such learning objectives it seems that essay questions are a better tool of assessment, as all can be categorized as informational topics. However, this is by no means to say that forced-choice questions fall short of assessing these objectives. On the contrary, if well constructed, forced-choice questions might prove to yield better results. At the same, essay questions, carelessly constructed, might turn into a bad too of assessment. As far as this study is concerned, it suffices to say that essay questions are a better tool method of assessment when it comes to informational topics.

METHODOLOGY

The research is based on a response questionnaire designed to be answered by the instructors involved in teaching two main courses in the Core Curriculum Program of the CBE. The questionnaire includes various variables related to the instructors, like age, gender, specialization (humanities or natural sciences). The population of the study consists of 500 full time instructors currently teaching in the CBE. A sample of 120 instructors from the population will be randomly chosen. The sample represents all the departments (about 11 departments) housed by the CBE, to make sure that both humanities and the natural sciences are represented.

Population and Sample

This research aims at exploring the factors determining the preferences of the instructors teaching the required courses in the Core Curriculum Program at the CBE with regard to types of tests (essay or forced-choice) and the rationale of their preference of one over the other when they assess their students. Of the six Core Curriculum requirements, we chose two courses to be the basis of this study: *Science Education* and *Kuwait and Development*. The results of the study are based on the response of 120 instructors of these two courses to a questionnaire and an interview. The sample was randomly chosen from a population consisting of 500 instructors involved in teaching Core Curriculum courses at the CBE. The sample represented three variables: gender, age and major. We believe that these two courses are fairly representative of core curriculum requirement in CBE for two reasons. First, the two courses fall within two major domains of knowledge in general education: the domain of the natural sciences and that of social studies. Consequently, results obtained from responses related to these two courses can be informative of the rest of the core curriculum courses. Second, the two courses involve several majors and departments in CBE, which offer and teach them. This means that the sample of the study; i.e., the instructors of the two courses come from several departments, and hence they can be fairly representative of the instructors involved in teaching the core curriculum requirements and electives.

Instrumentation

The participants responded to a 21-statement questionnaire designed to give informative answers to the research questions. 165 questionnaire sheets were distributed to members of staff in the CBE, of which, only 120 were completely filled in and used for this study. Each questionnaire started by inquiring about the gender, age and major (humanities or sciences). These variables will, later on, help in finding differences based on the independent variables.

Other than the independent questions, the questionnaire consisted of 21 statements, which used a scale of "agree, disagree, or don't know". The questions addressed four main categories compatible with the research questions of the study; namely, the types of questions used, the tests and their compatibility with the objectives of the course, factors determining the preference of one form of test over the other, and the level of understanding and knowledge the tests measure. In addition, the researchers interviewed 12 members of staff, asking them about their preferences when it comes to testing, and the factors determining their preferences. The members were selected randomly from both humanities and science major. Age was not requested during the interviews, yet the participants were from both genders. The answers from the interviews were manually analyzed.

The results of the 21 statements were analyzed via SPSS in search for mean, frequency and standard deviation. Significance was also investigated where the independent variables interact with the 21 statements and significance was calculated at $p \leq 0.05$.

RESULTS

The Questionnaire

The results show that the majority of teachers use both essay and forced-choice questions in the same test (78.4%). However, 18.6% percent only give tests which are exclusively essay type, and 28.4% are in favor of forced-choice exams only. However, high significance was found in a T-test by age and by gender showing that the younger female teachers disagree most with essay tests and agree most with forced-choice solely (0.001 and 0.040 respectively).

The first category of analysis consists of six statements, which are based on the ability of the exams to measure the learning objectives of the course and cover the material properly. Most teachers believe that their choice of one type over the other depends mainly on the objectives of the course being taught. Moreover, 78.4% of the instructors used forced-choice tests, which vary to address the course goals. Some instructors stated that they provide four or five questions and ask the students to answer three of four, or four of the five questions provided, as they feel it covers the objectives and gives students a choice of the test form they prefer. In a one-way ANOVA by gender, significance is found at 0.043 as male teachers design the classical multiple-choice method most. A two-way ANOVA by gender and age shows that the younger female teachers mostly disagree with the classical method (0.004). In addition, a three-way ANOVA by gender by age and by major shows high significance where p=0.044; it was found that the younger male teachers from the scientific departments are the most who use the classical type of testing.

The results also show that the majority of participant teachers do not feel that they are able to cover all the material required and cannot investigate details in forced-choice tests (54.9% and 59.8%). Yet, surprisingly, 59.8% believe that forced-choice testing is a suitable method to cover the aims of the course. Hence, many teachers do not believe that slight details are required and the memorization of all the information is important. What is important, as it seems, is the coverage of the course aims.

The second category, containing five statements, investigates the time required to correct essay and forced-choice exams, and the effort needed to put together a proper exam for each type. In terms of making an exam, half the members of staff participating in this study feel that forced-choice exams are easy to design, while the other half disagree. In a one-way ANOVA by gender, significance is found at p=0.043. This is because male teachers have mostly felt that they do not know how difficult forced-choice tests are compared to essay tests. 60.8% of the participants stated that they do not choose the test type "because it is easy to correct"; however, 37.3% choose forced-choice because they have huge classes, and the

correction would be very difficult. In a T-Test, significance was found by age by gender (0.040), where female teachers mostly used forced-choice to spend less time correcting exams. Many members of staff do not make essay tests because they feel they are "subjective", and 65.7% would like to avoid essay tests as they are time consuming.

The third and last category in the questionnaire consists of seven statements, and targets the knowledge tested and the reflection of understanding of students. Most teachers believe that essay tests reflect the students' knowledge better than forced-choice ones. 59.8% feel that students are more creative in essay test. In a one-way ANOVA by gender, significance was found at p=0.043, where the female participants mostly felt that they "don't know". A T-Test by age and by major reflected significance of p=0.047, as the older science majored participants agreed most that essay tests enable students to be more creative. Hence, many teachers use essay questions to ensure the students are aware of all the course aims, and encourage students to memorize all the information presented during the course, unlike forced-choice tests. 40.2% of the participants felt that forced-choice questions reflect better knowledge of the course content.

The Interview

The members of staff interviewed were asked which type of questions they preferred, and on what bases they made their choices. 7 out of 12 participants believed that any exam should have both types—essay and forced-choice. While 4 felt that forced-choice tests cover the needed information, and would reflect the students' full understanding of the course content; three of which were science major. Only one participant stated that essay questions are enough and a more practical method, as students would be able to answer the question in the form of a story, which is linked in terms of information and would be very creative.

Many teachers felt that even the design of the exam depended on the time available. Therefore, many teachers preferred to make essay tests as they are easier to design; yet 5 teachers stated that they choose to take more time making forced-choice exam because the correction would take less time. The majority of teachers who preferred to design forced-choice exams were science major.

When the participants were asked about the major concerns they have regarding the exam types, 50% of the participants did not have many forced-choice exam because students tend to cheat more. As the number of students reach 60 students in many cases, having a forced-choice based exam encourages cheating. On the other hand, 33.3% of the participants reflected that their concern over the ability of students to explain themselves clearly in essay exams. One member of staff added that "I feel like they can answer the questions, but could not express it clearly."

Almost all the participants (all but one humanity teacher) felt that students prefer forced-choice tests. 6 participants believed that students with lower GPA prefer forced-choice exams; in addition, some participants believed that many students feel less stressed knowing they have a forced-choice exam. Although 6 teachers stated that forced-choice exams could only be used when the content of the subject is suitable for that type, the same teachers stated that they use forced-choice exams because the students prefer that type..

Most participants in the interview asked to see the results of this study, and some asked for a study that looks closer at the students' perception regard testing types. They sensed that a teachers' point of view would be concentrated more on the academic outcomes, while the students might consider the way they study and the amount going to be studied more.

DISCUSSION

This study aimed at investigating the instructors' preferences in designing their tests for core curriculum courses. Four main ideas were the focus of the study: the type of questions

chosen, the factors determining their choices, the compatibility of the tests with the course learning objectives, and the levels of thinking the test addressed. The answers obtained are discussed from two points of view: statistics and literature.

Clearly, the results have shown that the majority of instructors (78.4%) prefer to design tests that combine forced-choice and essay questions. Yet, of the instructors who don't prefer to combine both types of questions in their tests, the majority (60.4%) favors forced-choice questions over essay questions, where male participants significantly prefer forced-choice tests. However, though the majority of instructors use forced-choice type, the statistics also revealed that they do not believe that the forced-choice tests could cover all the aspects of the curriculum needed by the instructor. A close reading of these results seems to reveal that the factors determining the choice of one method of assessment over another are peripheral to the process and steps stipulated in designing classroom tests. That is, from these results and the results of the interviews with instructors, the time taken in grading the test and the class size are crucial determinants in their choices. This result is consistent with the results obtained from other questions in the questionnaire.

The instructors who did not favor the forced-choice questions rationalized their choice by referring to the fact that the large number of students in their classes makes cheating a very easy task. In addition, a few instructors disfavored the forced-choice as a method of assessment because the design of forced-choice items is time consuming. It is worth noticing here that on both ends of the spectrum favoring one method over the other is underlined by determinants external to the curriculum content and is based mainly on the time consumed (in design or scoring) and class size. Furthermore, as Marzano (2000) has indicated, forced-choice questions are very difficult to construct because the viable distractors need to be easily identified as "wrong" by the students aware of the course content, and reasonable to the students who rely on "educated guess". On the other hand, essay questions are easy to design, as one question may cover a complete chapter, but would take a much longer time to correct.

The majority of instructors who favored forced-choice tests believe that this type is a suitable method of assessment for their courses since knowledge, comprehension and retrieval of information are integral to the learning objectives of the courses. As such, they were of the opinion that forced-choice items, particularly conventional multiple-choice question, are better tools for assessment as careful design of the correct choice and plausible distracters would show deep understanding of the material on the students' part. One of the conclusions that can be gleaned from this discussion is that instructors are aware of the importance of the compatibility of the tests with the learning objectives that these tests assess. Favoring forced-choice items over essay item does not undermine this awareness. As we have seen in the literature review, measurement experts give forced-choice forms the score of M for informational topics, which means that forced-choice forms can be used to asses these skills and processes, but better assessments (basically essay forms) may be used.

Some instructors (40.2%) prefer to use forced-choice or essay exams, where students are asked to 'choose three out of four' to answer. This offer of choice is justified by the instructors as aiming at helping students by providing options, yet most teachers who use this method of testing believe the questions that fall under one category should cover one aim. Therefore, the students would be given the chance to answer the questions that are more appealing or easier to answer. It appears that the students' preference and liking has become one of the fundamentals of test design.

Overall, the instructors' preference for forced-choice questions is based on three determining factors. One of the main factors was the fact that forced-choice tests are easier and less time consuming to grade than essay forms. The second factor is students' preference for forced-choice questions as these questions jog their memories for the details. In other words, the majority of instructors believe that the students prefer forced-choice tests and, therefore, they do not mind catering for the students' needs as well. The third and final factor is the instructors' conviction that essay tests might be unfairly corrected, and are subject to bias.

This idea was presented previously by Valenti *et al.* (2003) who were looking for an automated method of essay grading, and a large number of researchers seem to agree (Grondlund, 1985; Page 1994, 1996 among others).

On the other hand, the instructors who favored essay tests as a method of assessment in their courses seem to work with different determining factors. Basically, they seem to believe that essay tests can measure more aspects of the students' knowledge and understanding. That is to say, they viewed essay questions as a better method when it comes to the skills and processes the students should master. As for the subjectivity the essay questions are prone to in grading and scoring, the instructors believe that this pitfall can be easily avoided once the instructor sets a criterion beforehand on the amount of knowledge and type of information required from that specific question.

The last research question in this study addressed the levels of understanding and orders of thinking measured according to Bloom's Taxonomy (1956). As we have seen in the tables of specification for the learning objectives of the two courses in question, all the objectives address the lower-order thinking skills: knowledge, comprehension and retrieval of information. Hence, the data collected showed that the majority of teachers are only testing the first two levels of understanding: knowledge and comprehension. Very little attention has been given to Marzano's (2000) what you can do with what you have learned? Rather, the focus is merely on what the students have learned, either in detail or in a general sense. The majority of teachers participating in this study have only discussed the recall of information, the assurance of memorization and time consumption. The only time "understanding" was mentioned, was actually used to refer to the ability to translate information in the students' own words and use the information acquired in the correct examination space. The fact that these tests fall short of measuring higher-order thinking skills provides commentary not on the inability of instructors to design tests that take students up the ladder of the thinking skills, as much as on the deficiencies in the design of the course curriculum and the program at large. The courses' learning objectives are limited to the first two levels of thinking, which, unfortunately, do not provide students with opportunities to explore higher domains of thinking. It inevitably follows that the tests mirror the content and the learning objectives.

It is worth noting that the independent variables of age, gender and major had very slight effect on the outcomes of the current study. Although male students have shown significant preference of forced-choice test, it is evident that the majority of teachers prefer that specific type. In addition, while science-majored teachers significantly prefer essay type tests more than humanity-majored ones, this did not change their forced-choice based tests. This is due, as mentioned, to the fact that extrinsic factors determines the test design more that factors integral to the processes and steps of test design, like the class size and the time consumed in designing and scoring the tests

CONCLUSION

This study aimed at exploring factors that determine the instructors' preferences in designing tests for the Core Curriculum required courses in CBE, Kuwait. For this purpose, the study posed four interchangeable questions. The first question tries to explore the type of tests (forced-choice or essay questions) instructors favor/disfavor. The results showed that most instructors favor tests which combine both types. Of the instructors who design tests with either forced-choice items or essay items, the majority favor forced-choice over essay questions. The answer to the second question reveals that the factors determining the choice of one type over the other are basically extrinsic to the processes and steps of test design, like the class size and the time consumed in designing and scoring the tests. Of course, these factors are prioritized for the instructors because of the special circumstances related to CBE's policies of accepting a large number of students without strategically planning for

proper teacher-student ratio. Thus, the compromises made at the level of the College administration stipulates further compromises made by the instructors who try to deal with classes packed with students. Under these circumstances, instructors are inclined to disregard factors intrinsic to designing the test like the ability of the tests to measure the learning objectives, and focus, instead, on ways to deal with the problems created by the CBE's policies of admission. Needless to say, this vicious circle would jeopardize the quality of education and the academic accreditation that the College is seeking. As for whether the tests designed by the Core Curriculum instructors address the various levels of thinking, the study showed that the tests are limited to the lower-order thinking skills: knowledge, comprehension and retrieval of information. The question items, whether forced-choice or essay, don't venture into higher domains of thinking, like application, analysis, synthesis and evaluation. However, the problem here is attributed not to the test designers as much as in to the course designer who limited the learning objective to the lower levels of thinking. Again, this fact puts at stake the implementations of quality education and academic accreditation project.

RECOMMENDATIONS

According the discussion above, the researchers recommend the following:

1. To improve the quality of tests designed for the Core Curriculum program, the researchers recommend that the CBE administration reconsider its policies of admission and hence redress the imbalance of the teacher-student ratio. This step might shift the focus from the factors external to the test design, like the time consumed in grading and the class size, to factors integral to the test design, like the learning objectives and the skills and processes that tests should address.

2. The researchers recommend that mandatory training courses and workshops on designing classroom tests be held for all instructors. Such training courses have the virtue of introducing the instructors to the steps and methods of designing tests.

3. The researchers recommend that the CBS undertake a comprehensive and broad review of its programs in order to integrate objectives address that various levels of thinking. This step is crucial not only in improving the quality of tests and their ability to venture into high domains of thinking, but also in improving the quality of education

REFERENCE

Abu 'Allam, Rajaa (1987). *Measuring and evaluating performance* [Arabic]. Kuwait: Al-Qalam Publishing House.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: McKay.

Cameron, J. (1996). "Reinforcement, reward and intrinsic motivation." *Review of Educational Research*, 64(3), 363-423. http://dx.doi.org/10.3102/00346543064003363

Carr, Nathan T. (2015*). Designing and analyzing language tests*. Oxford: Oxford UP.

Gipps, Caroline V. (2012). *Beyond testing: Towards a theory of educational assessment*. London: Rouledge.

Gronlund, N.E. (1995). *Measurement and evaluation in teaching*. New York, Macmillan.

Haladyna, T. M. (1994). *Developing and validating multiple-choice test items*. Hillscade, NJ: Erlbaum.

Jaber, Abdulhamid J. (1993). *Readings in teaching thinking skills and the curriculum*. [Arabic]. Publications of the Center for Human Resource Development. Cairo:Al-Nahda Al-Arabiyya Publishing House.

Marzano, Robert J. (2000). *Transforming classroom grading*. Virginia: Association for Supervision and Curriculum Development.

Madaus, G. (1988). "The influence of testing on the curriculum" in Tanner L. N. (Ed.) *Critical Issues in Curriculum*, 87th Yearbook of NSSE Part 1, Chicago IL, University of Chicago Press.

Mindes, Gayle, et al. (1997). *Assessing Young Children*. Boston: Delmar Publishers.

Muqadam, Abdulhafith (2003). *Modern trends in evaluation: The perspective of quality education and academic accreditation*.[Arabic]. Riyad.

Obaidat, Suleiman A. (1988). *Measurement and evaluation in education*. [Arabic]. Amman: Jordan UP.

Owdah, Ahmad S. (1993). *Measurement and evaluation in education* [Arabic]. Amman: Al-Amal Publish House.

Stiggins, R. J. (1994). *Student-centered classroom assessment*. New York: Merrill.

Valenti, S.; Neri, F. & A. Cucchiarelli (2003). "An Overview of Current Research on Automated Essay Grading". *Journal of Education Technology Information*, Vol. 2, 2003.

Wortham, Sue Clark. (2002). *Tests and Measurement in Early Childhood Education*. New York: Macmillan.