

ISSN 1799-2591

Theory and Practice in Language Studies, Vol. 7, No. 8, pp. 597-607, August 2017

DOI: <http://dx.doi.org/10.17507/tpls.0708.01>

# Authentic Assessment: Evaluating the Saudi EFL Tertiary Examination System

Fahad Saleh Suleiman Alfallaj

College of Science and Arts, Methnab, Qassim University, Saudi Arabia

Arif Ahmed Mohammed Hassan Al-Ahdal

College of Science and Arts, Methnab, Qassim University, Saudi Arabia

**Abstract**—As early as 1983, Rossi propounded that one of the issues of particular interest and development within the foreign and second language teaching profession is that of proficiency testing or the evaluation of a learner's level of linguistic and communicative competence. This still holds true. On the contrary our pilot study using the Question Paper Evaluating Checklists (included in this paper) indicates that all is not right with the designing of EFL question papers in Saudi Arabia though EFL assessment patterns in the KSA have undergone much change from the time that English was first introduced into the curriculum as a compulsory foreign language. It is the demand of time that evaluation patterns be evaluated on the touchstone of latest research and their relationship with classroom practices be established. This will help the learner-teacher combine to plug the loopholes in language training. In other words, we have to realise as educators that good assessment forms the basis of a wealth of learner information that has direct and indirect ramifications on curriculum and pedagogy. Hence the need to study this aspect of EFL in the light of modern literatures in order come up with constructive recommendations.

**Index Terms**—EFL, question papers, assessment, testing, evaluation

## I. INTRODUCTION

The aim of a language test is to elicit such language use responses from the learners as may enable us to arrive at conclusions to predict the learners' ability to use language in contexts outside the test situation alone. Further, in EFL contexts such as in KSA, future employment depending upon the learners' grading, certification, selection and finally, placement remains the larger motivation for enrolling in foreign (English in this study) language courses. This makes it all the more imperative for the tests to be designed and used to gauge these specific parameters. Moreover, the testers would want to show (1) that the components that they taught in the courses offered also correspond to the items included in the test; and (2) that the learners are familiar with the test tasks as they replicate the learning activities they actually engaged in during classroom learning. In short, the language test will help us establish a specific relationship between classroom sessions, test expectations (of the testers) and real life language use by the learners.

The current paper proposes to focus specifically on the testing of the four English language skills under the Degree Programme called English Language and Translation at Qassim University, KSA. It is often seen in EFL testing situations that whereas Reading and Writing are assessed by assigning comprehension and free writing tasks, more often than not, in a test of Speaking-Listening skills, ease overtakes principle and practical considerations defeat desire. The end result is that we administer a written test where our aim is to test the speaking proficiency of the learners. This is also observed in the KSA, where the closest the assessment comes to testing speaking is in shifting the focus to translation. We forget the old adage that 'Translation is an art' and ought to have a limited place in language learning as such. In fact, the rigours of the course should aim to prepare the learners to acquire enough skill to apply themselves to translation: It should be the outcome of the course rather than the course itself.

This paper also attempts to develop a checklist for the evaluation of EFL test papers.

The syllabus for EFL, a course known as Bachelor's Degree in English Language and Translation with a Major in English and Translation, lists the following components (not necessarily in this order; sometimes a topic is spread over more than one unit):

1. English Grammar
2. Reading and Vocabulary Building
3. Listening and Speaking
4. Pronunciation
5. Writing
6. CALL
7. Translation Theories
8. Instantaneous Translation
9. Literary Translation

This is an eight term course in which at the end of the fourth term learners have to choose between Literature and Linguistics in addition to Language which is a compulsory component. For reasons of space and time, we have limited out investigation to the Literature component to the exclusion of Linguistics which can be an exclusive area of study.

Clearly the philosophy behind this syllabus design is to treat language proficiency as a multi-dimensional ability as enumerated by Oller (1983) consisting in 'a number of interrelated specific abilities as well as general ability or set of general abilities or procedures'.

In line with this thinking, the Course Prospectus spells out the goals of the programme. Broadly stated, these are:

1. Training and preparing the learners to be translators and interpreters.
2. Improving the linguistic skills of KSA learners in Arabic and English.
3. Enabling the learners to appreciate diverse cultures and languages.
4. Helping the learners enhance their oral and written communication skills.
5. Developing learners' critical thinking skills and ability to produce logical and well-structured arguments.

In short, the programme aims to turn out language enabled people who have the requisite language tools to succeed in any field, whether it be law, politics, business, or any other situation in international work culture.

The aim of this paper is:

- i. to study whether and how far the Achievement Test mechanism of the University is able to evaluate the proficiency achieved by the learners' as a result of the programme;
- ii. if the findings show that there are pitfalls in the Achievement Test system, then suggest ways and means to rectify them.

## II. RESEARCH PROBLEM AND SIGNIFICANCE OF THE STUDY

Assessment patterns in KSA have not seen a sea change ever since English was introduced as a compulsory course in 1958. By virtue of their poor dynamism, it may be assumed that the assessment measures currently in vogue in KSA do not correctly assess the learners' potential to acquire new skills, and develop lateral thinking that fosters problem solving strategies in new and unfamiliar situations. By the latter we mean the real world of language use that EFL speakers encounter in higher education, research or work environments. The EFL papers lean heavily on testing the recall capacity of the learners. Instead of being an aid to learning, testing methods in EFL are indicators of learning alone. These do not test the learners' ability to interconnect ideas and use learned inputs to create new outputs. On the touchstone of variety in the design and content of the assessment tool(s) too, these papers fall flat on the face. In the long run, assessment tools are valid and useful only if they allow adaptation of teaching to student needs. This study is a significant step in the direction of remodelling the assessment patterns in EFL to tackle the issues highlighted here.

## III. RESEARCH OBJECTIVES

The study has the following points of reference as its objectives:

1. To highlight the shortcomings of the current testing patterns in EFL with special reference to concurrence between classroom learning and test content and pattern.
2. To provide guidelines on designing EFL question papers that have the following elements:
  - i. Coherence
  - ii. Effectiveness
  - iii. Inputs to assess declarative knowledge
  - iv. Inputs to assess procedural knowledge

## IV. RESEARCH QUESTIONS

1. Are the current EFL test papers made up of the items of highest priority for testing with common usage as the benchmark?
2. Is there duplication in the test items, e.g. more than one question testing the same skill?
3. Are the tests using optimum techniques for testing certain topics? For instance, some language elements lend themselves to dictation, others to completion, others to multiple choice, etc. Does the paper pay due attention to these factors?
4. Are a variety of techniques being used within one test to make it more interesting?
5. Does the test paper use the same language as used during the instruction of an item?
6. Do the learners feel they were given a fair opportunity by the questions to bring out their best?

Questions 1-3 will be answered by comparing the Qassim University EFL question papers with the MET sample (enumerated upon later in this study). A fifteen item questionnaire will be administered to collect primary data to answer Questions 4-6.

### Testing, Assessment or Evaluation?

All the three terms are used to measure learners' progress in the light of the goals and objectives stated in the policy that lies behind inclusion of a subject in the curriculum. However, these also indicate different things at a micro level. Testing measures the short term curricular achievement by checking the level of skill reached. Evaluation takes into

account the result of testing to make judgements on the criteria and evidence. Finally, Assessment is a process spread over a longer duration of time and is based upon consideration of learners' knowledge, skill acquisition, recording and scoring of these in measurable terms with the larger aim of improving the process of learning.

#### V. LITERATURE REVIEW: WHY TEST THE TESTS?

Summing up developments in language assessment over a period of many previous decades, So et al. (2015) assert that contrary to the finding that real world language use is simulated only by tasks that involve multiple language skills, current testing mechanisms measure one of the four skills at a time. For reasons of practicality, assessments do not integrate all the language skills. The answer to devising such tests might lie in the available literature: each offering a bit that can contribute to the whole.

In a discussion of steps that may aid in improving teachers' assessment of EFL learners, Cumming (2009) talks at length about the significance of teacher development. This finding is relevant to the current study also as it says that classroom based assessment is a factor of educating the practicing teachers to 'appreciate' and 'use' the vast assessment related information that has accumulated in the recent times. Further that teacher-student interaction is central to classroom learning as it inevitably involves much covert as well as formal assessment of the learners' performance. Thus, it sheds light on the responsibility that falls on the teachers when they opt for Formative Assessment discussed in the next study quoted here.

Quoting Black and William (1998), Xu (2009) says that conventional testing methods are plagued by the 'teaching-to-the-test' phenomenon among other problems. To rid the EFL system of this negative effect, Chinese TESOL educators have started favouring an alternate means of assessment. This is Formative Assessment. This enables the teacher to plan the instruction content and pedagogy based upon the feedback obtained from the learners' performance in the FA. However, one big challenge here is the effective implementation of the system.

NOTE: Formative Assessment (FA) is based on short term learning and can use Multiple Choice Questions (MCQs), fill-ups, true/false, or other objective testing patterns to check the students' learning.

Need for in-service teacher education, an exclusive area of study by itself, in improving testing patterns is explored by Troudi (2009) who sought teacher feedback based upon an open ended questionnaire. This study has special relevance to this work as it was carried in a similar language setting: it is a study of EFL assessment in the UAE and Kuwait, where the mother tongue is Arabic. The results showed that the factors affecting the teachers' views on assessment were: their knowledge of the field of language learning and teaching, and by the prevalent social milieu and socio-political factors governing their employment status. In other words, the teachers had little, if any, role to play in assessment because of the 'top down managerial approach' to education, concern for validity and quality assurance in the programmes.

Looking at early history of research into language assessment, what Rosenbaum (1971) says about the 'unpleasant' task of testing is as true today as it was then: Since tests determine marks, and marks result in student attitudes toward teachers as well as subjects, it follows that tests play a crucial role in the student-teacher relationship. To be sure, good testing is based on good teaching; in both activities knowledge of the subject is as important as professional skill. However, in testing more than in any other phase of teaching, lack of know-how will be of consequence. Further that teachers in training and beginning teachers know very little about testing. Experienced teachers when faced with poor test results are some- times apt to blame students and not themselves. "He just didn't study." But often the test itself, the method of administration or that of marking is at fault.

In the following decade, however, empirical studies clearly demonstrated that the kind of test tasks used can affect test performance as much as the abilities we want to measure. Pioneers in this promulgation were Bachman & Palmer (1981, 1982, 1988); and Clifford, (1981) to name a few.

Studies by Alderson & Urquhart (1985) and Erickson & Molloy (1983) among others demonstrated that the topics of test tasks can affect performance. Results of these studies have stimulated a renewed interest in the investigation of test content.

As quoted in Bachman (1991), Alderson and colleagues (Alderson, 1986, 1990; Alderson & Lukmani, 1986; Alderson, Henning, & Lukmani, 1987) have been investigating:

(a) the extent to which "experts" agree in their judgments about what specific skills EFL reading test items measure, and at what levels, and

(b) whether these expert judgments about ability levels are related to the difficulty of items. Their results indicate first, that these experts, who included test designers assessing the content of their own tests, do not agree and, second, that there is virtually no relationship between judgments of the levels of ability tested and empirical item difficulty.

According to Bachman (1991), test designers and experts in the field disagree about what language tests measure, and neither the designers nor the experts have a clear sense of the levels of ability measured by their tests.

The 1980s in Language Testing were a decade of "communicative" testing. Two branches of communicative approaches to language testing came into existence: One traces its roots to the Canale/Swain framework of communicative competence (Canale, 1983; Canale & Swain, 1980). The other is characterised by the Test of English for Educational Purposes (Associated Examining Board, 1987; Weir, 1983), the Ontario Test of English as a Second Language (Wesche et al., 1987), and the International English Language Testing Service (e.g., Alderson, 1988b;

Alderson, Foulkes, Clapham, & Ingram, 1990; Criper & Davies, 1988; Seaton, 1983) that grew out of the English for specific purposes tradition.

According to Bachman (1991), the following are the features of Communicative Language Testing:

1. An "information gap," requiring test takers to process complementary information through the use of multiple sources of input. Test takers, for example, might be required to perform a writing task that is based on input from both a short recorded lecture and a reading passage on the same topic.
2. Task dependency, with tasks in one section of the test building upon the content of earlier sections, including the test taker's answers to those sections.
3. Integration of test tasks and content within a given domain of discourse.
4. Finally, communicative tests attempt to measure a much broader range of language abilities- including knowledge of cohesion, functions, and sociolinguistic appropriateness- than did earlier tests, which tended to focus on the formal aspects of language-grammar, vocabulary, and pronunciation.

## VI. RESEARCH METHODOLOGY

The ultimate aim of all language training is to bring the learners up to a level of proficiency where they can use the language successfully as a medium of communication. Every year students graduating from various universities in KSA aspire to proceed for higher education in Europe, Great Britain and the USA. An equally large number wishes to find employment in the western world. In either case, they are required to appear for standard English proficiency tests to determine their language ability.

In the KSA English is taught from grade four onwards as a compulsory foreign language. However, in informal interviews, learners have variously confessed their disappointment with the pedagogy, syllabus content, and even testing methods, blaming either all or a combination of these for their less than desirable proficiency. This feedback will also be verified formally by administering a questionnaire (Appendix 01) as discussed later here.

Much is wrong overtly and covertly with the test papers at QU. To start with the most worrisome issue, these papers do not ever cover the complete prescribed syllabus. Thus this kind of assessment does not truly reflect the learning. Neither does it test the students' understanding or their ability to apply this understanding.

Then, the assessment rubric or instructions at the beginning of the paper or each question fails to specifically state what the students are expected to do. How can we expect the students to do well when in the first place, they are not even aware of the criteria for success? In fact, anything that is to be graded or assessed, should have a criterion already given to the students even before they are taught it in the class.

In foreign language learning, short term memory is more actively involved than long term memory. However, the EFL papers in QU run contrary to this logic. Questions that begin on one page are often carried over to the next page. By the time the students read through the entire question, their memory has already relegated to the background what they read on the previous page. Consequently, they either waste precious time flipping the pages or end up answering what they have not even been asked!

In an assessment situation, it is important for students to have a clear picture of the weightage assigned to each question so that they can prioritise their answers. This is even truer for foreign language assessment where the learners' proficiency is only borderline. However, the QU papers do not state the maximum marks allotted to each question, and worse, sometimes the marks allotted do not add up to the total weightage of the paper.

Further, there have been instances of students failing to distinguish between the instruction section of the question and the question itself. The reason is that the two sections mostly run in continuation with no change of font or space break to differentiate between them. Moreover, these huge chunks of printed material with no variation in font style or size appear intimidating to the already stressed out student of foreign language, adding in multiples to his/her confusion and loss.

Consequently, this study will follow a multi-pronged approach:

- (i) Current EFL testing instruments of Qassim University will be compared with the MET (Michigan English Test) which is a standardised test to conclude whether the KSA EFL testing pattern conforms to those globally recognised.
- (ii) The learners (80) enrolled for the two EFL courses will be requested to attempt the **free online sample** of the test and submit the self-assessed scores for compilation and analysis. These results will be used to draw correlation with their performance in tests of these language items administered in due course at the University.
- (iii) A fifteen item questionnaire will be administered to forty EFL learners each in the two EFL courses offered at Qassim University, viz, Language and Literature, and Literature and Linguistics, to get student feedback on EFL question papers at the University.

## VII. ABOUT THE MICHIGAN ENGLISH TEST

The official site of MET [www.cambridgemichigan.org](http://www.cambridgemichigan.org) clearly states that the MET is a standardised test (that is one that has been designed as a result of research) that tests the high-beginner to low-advanced EFL learner. It is administered in many countries around the world to gauge the English language proficiency in 'social, educational, and workplace contexts'. Clearly, the test covers practically all possible language use situations that the Saudi EFL learner

is likely to encounter in real life. This test is regularly updated to keep pace with ongoing language and assessment research and is so designed as to help the test takers demonstrate their language proficiency rather than discover the areas they lag behind in.

The MET is variously used to judge proficiency at the end of the course, and in employment situations, such as, selection and promotion where language proficiency is desirable. MELAB, offered by the CaMLA is one of the Michigan Battery Tests and is used for admissions to academic courses in many countries in the west.

As mentioned before in this paper, the current study proposes to use only the **free online sample** of the test as the baseline to compare the learners' earlier performance in similar items. The coefficient of correlation so arrived at will enable us to determine the standard of the current EFL testing machinery at Qassim University.

We have chosen MET even though it tests learners primarily on English Language Skills whereas at Qassim University they are also tested for content courses. This is based on the premise that foreign language learners may receive lower scores on content area assessments administered in English than they would if they took the same tests in a language in which they were proficient. This is also the reason why international English language tests too limit themselves to the skills.

#### VIII. FINDINGS, DATA ANALYSIS AND INTERPRETATION

Language proficiency is a dynamic ability, ie, one that can improve or worsen over time and practice. A language score in say, Grammar, in the first term of the course need not be valid by the time the learners reach the last term and undergo the Exit Exam. Testing one language component in a particular term and then forgetting all about it in the other terms does not in any way establish the learners' continued proficiency in that component. At Qassim University, the syllabus for English is designed in the following manner:

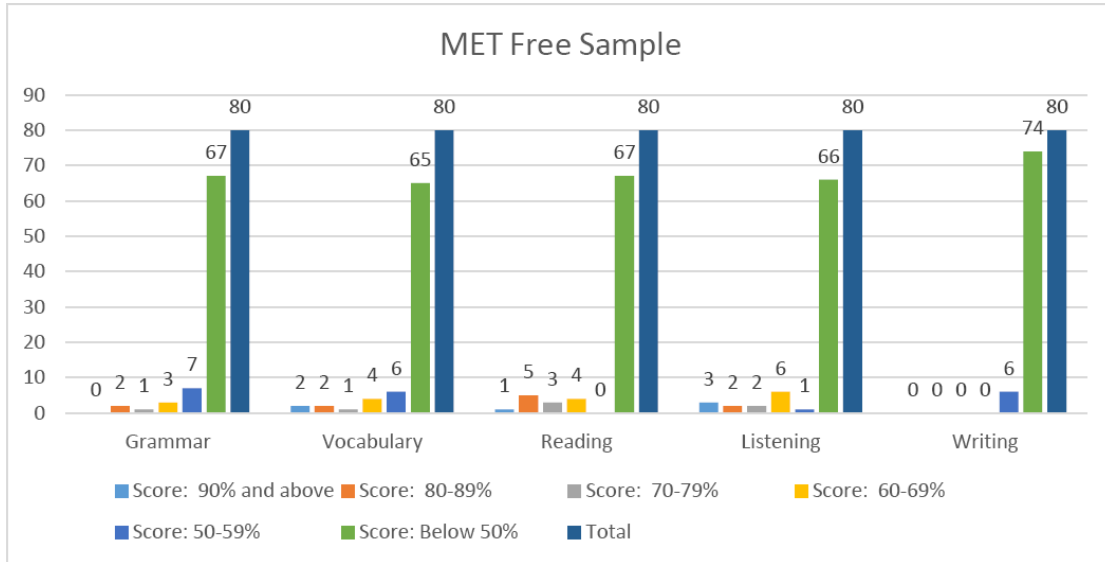
TERM	PAPER	PART	UNITS
	English Grammar & Sentence Writing-1		2 units
	Listening & Speaking	1	2 units
	Reading & Vocabulary Building	1	3 units
	Pronunciation		3 units
	English Writing	1	2 units
Term 2:			
	English Grammar 2		2 units
	Listening & Speaking	2	2 units
	Reading & Vocabulary Building	2	3 units
Term 3:			
	CALL		3 units
	English Grammar	3	3 units
	Listening & Speaking	3	2 units
	Reading & Vocabulary Building	3	3 units
Term 4:			
Term 5:			
	Translation theories		3 units
Term 6:			
	Machine Translation		3 units
Term 7:			
	Terminology and lexicography		3 units
Term 8:			
	Instantaneous Translation		3 units
	Literary Translation		3 units
	Issues and Problems in Translation		3 units

Consequently, the corresponding language items are tested at the end of each term. Thus, by the end of the eighth term, a long gap is established between the learners and language components covered in the first four terms. This is so as with the commencement of the fifth term, they are required to opt for specialization. What is forgotten is the fact that English for these (or any) learners in KSA is a foreign language and the test score of any component tested earlier may not hold good any longer.

This assumption is proved right by the learners' score in the comprehensive MET sample test where they are tested for several language skills within the same test paper. Eighty learners who would be taking their University Exit Exam within a few months were asked to undergo the MET. Out of these eighty learners who individually underwent the Free MET sample test available online, the component-wise scores were as follows:

TABLE 1  
MET FREE SAMPLE SCORES:

Component	Score 90% and above	Score 80-89%	Score 70-79%	Score 60-69%	Score 50-59%	Score Below 50%	Total
Grammar	0	2	1	3	7	67	80
Vocabulary	2	2	1	4	6	65	80
Reading	1	5	3	4	0	67	80
Listening	3	2	2	6	1	66	80
Writing	0	0	0	0	6	74	80

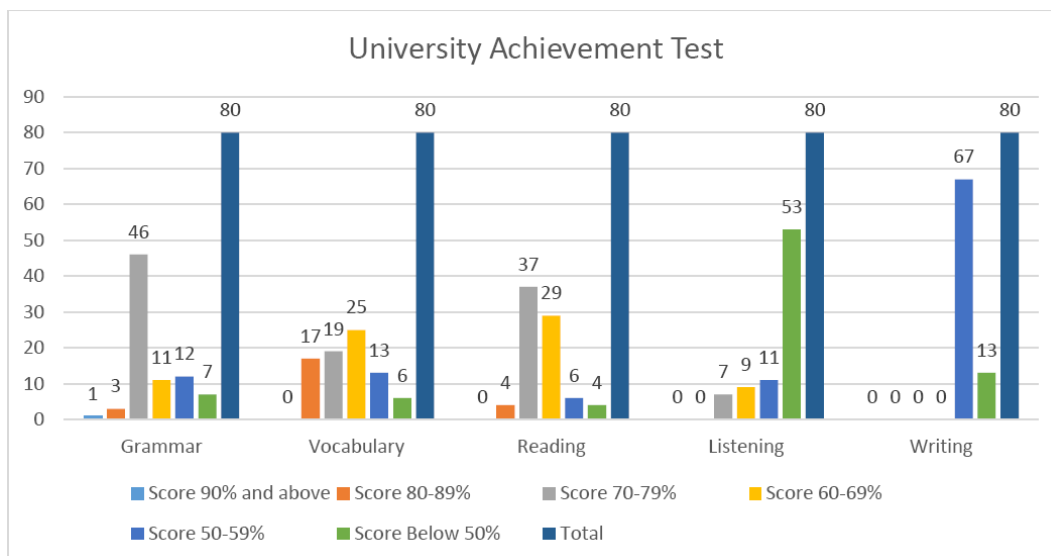


Clearly, in a term where the learners are asked to focus strictly on translation, other components of the language were not given sufficient classroom time and the result is as could be expected.

However, these results are rather contrary to the learners' in-house Achievement Test at the University. We reproduce here their scores.

TABLE 2  
UNIVERSITY ACHIEVEMENT TEST SCORES:

Component	Score 90% and above	Score 80-89%	Score 70-79%	Score 60-69%	Score 50-59%	Score Below 50%	Total
Grammar	1	3	46	11	12	7	80
Vocabulary	0	17	19	25	13	6	80
Reading	0	4	37	29	6	4	80
Listening	0	0	7	9	11	53	80
Writing	0	0	0	0	67	13	80



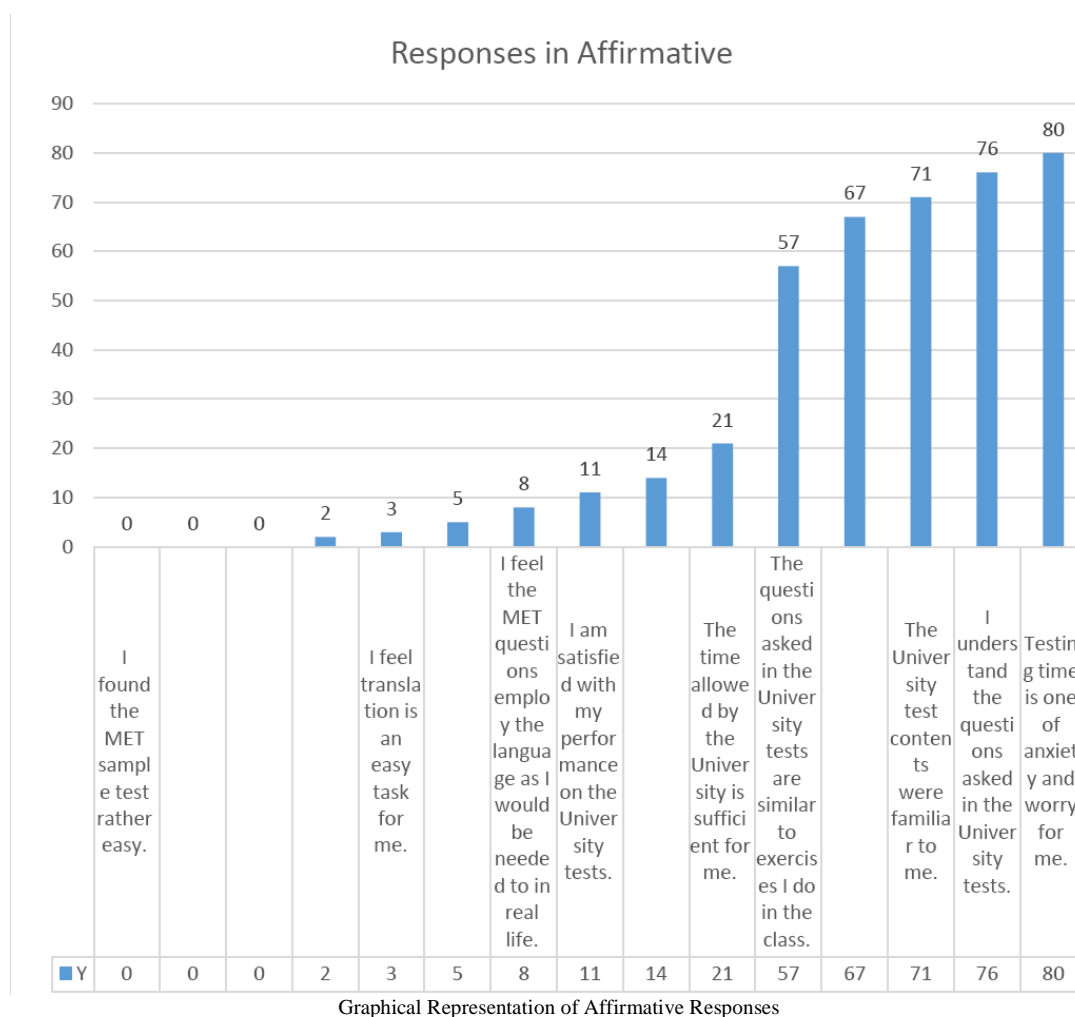
Checking the results at a glance, we observe that whereas in MET the tendency of the data (comprising learner output) is more on lower percentage of scores, the trend dramatically reverses in the University Achievement Test where the scores show a far better leaning with more concentration at the middle.

This comparison clearly establishes the fact that the respondents are not prepared to succeed in the internationally recognised proficiency tests even if they are moderately comfortable with the pattern and content of the UETs: the reasons are many as are discussed later here in the Recommendations section. In the Grammar component, only seven participants scored poorly (below 50%) in the UET but in the MET, this number went up to sixty seven. In Vocabulary section, the scores are similar, with sixty five participants still scoring less than fifty percent marks but in UET they seventy four participants scored between eighty nine and fifty percent marks. This would indicate one of the two possibilities: i. the UET tests simple vocabulary; ii. the participants tend to learn the vocabulary to be tested in the University exam. In the Reading test, again sixty seven participants scored less than fifty percent marks but a large number of this same group scored between sixty and seventy nine percent marks. The outcomes are much the same in Listening test with sixty six participants hovering below the fifty percent mark in MET but a close fifty three participants in UET also did poorly with scoring below fifty percent marks. This reinforces our assumption that the oral-aural skills are ignored in the University curriculum. Finally in the test of Writing seventy four participants scoring below fifty percent in MET but only thirteen doing that poorly in UET.

Let us correlate this outcome with the learners' response (Table 3) to the Questionnaire (Sample Appendix 01).

TABLE 3  
QUESTIONNAIRE RESPONSES:

S. NO.	QUESTION	Y	N	TOTAL
1	I understand the questions asked in the University tests.	76	4	80
2	The time allowed by the University is sufficient for me.	21	59	80
3	I am aware that for employment I may have to take one of the International Proficiency Tests for English.	2	78	80
4	Testing time is one of anxiety and worry for me.	80	0	80
5	I am satisfied with my performance on the University tests.	11	69	80
6	I found the MET sample test rather easy.	0	80	80
7	The questions asked in the University tests are similar to exercises I do in the class.	57	23	80
8	My performance would have been better had the University test components been different.	14	66	80
9	I feel translation is an easy task for me.	3	77	80
10	For my aspirations from the EFL course, I find the tests right and closely related to real life situations I might use the language in.	0	80	80
11	The University test contents were familiar to me.	71	9	80
12	I was able to tackle the MET sample test questions because I had studied these components in my course.	5	75	80
13	My teacher taught me in the class the way questions are asked in the MET sample test paper.	67	13	80
14	I feel the MET questions employ the language as I would be needed to in real life.	8	72	80
15	With my knowledge of English, I am confident of scoring well on any proficiency test for English.	0	80	80



### IX. RESPONSE ANALYSIS

An overwhelming ninety five percent of the learners are familiar with the language used to ask questions in the University Achievement Test (henceforth UAT) but seventy three percent of them would have liked more time to write their answers. This goes to show that though they comprehend the questions well, they feel hampered by poor writing skills.

Ninety seven percent of the respondents are not aware of the need to undergo International Proficiency Tests in English for employment outside of KSA. Needless to say, motivation is a prominent deciding factor in foreign language learning. Ignorance about future needs passively eliminates motivation for these learners.

All the respondents uniformly reported feeling anxious for the UETs. This can be supposed to stem from lack of confidence which is a direct outcome of insufficient classroom drill. Eighty six per cent of the respondents also reported being not satisfied with their performance in the UETs.

In a finding that need not be surprising, the respondents though scoring moderately on the UET unanimously found the standard MET rather difficult. This is a significant finding for the 'washback' effect as it goes on to prove that the test components of the UET are not up to the International proficiency expectancy level.

Seventy one percent as opposed to twenty seven percent of the participants reported familiarity with the questions asked in the MET. This finding is also corroborated by response to Q1, Q8 and Q11 which seek to find whether or not the learners understood the UET question tasks; if the respondents would rather have had different set of tasks asked of them in the UET; and the learners' familiarity with the test contents.

The positive response of eighty seven percent of the learners to their inability to do the task related to translation, indicates poor proficiency which may be a factor of dearth of vocabulary as is also apparent in both MET and UET performances.

With good exposure to English films and music, the response of the learners to Q10 is educative for the teachers. Hundred percent of the respondents are aware that their training in EFL is not conducive to real life language use situations. Clearly, it is time to modify the EFL curriculum at Qassim University.



To the contrary, ninety three percent of the participants reported unfamiliarity with the MET test components (Q12) and testing pattern (Q13). Needless to say, MET is one of the standard English Proficiency Tests to qualify which may a prerequisite for employment outside of KSA. Responses to Q14 and Q15, wherein ninety one percent of the participants expressed awareness that the test items of MET appeared closer to their real life needs, and all eighty unanimously showed awareness of their poor English proficiency, are both very encouraging outcomes for the policy makers and teachers alike. With their current level of awareness, a change in the University curriculum and assessment design will be welcomed by the student community making it much easier to implement.

#### X. RECOMMENDATIONS

EFL testing at Qassim University can serve the interests of the learners better by modifying test machinery in keeping with the times. Here are our recommendations:

1. Exclusively testing one or a few language components does not ensure evaluation of other components. Therefore, after the fourth term too Grammar, Vocabulary, Listening, Speaking and Writing should find a place in the exams.
2. Not only this, classroom activity should also include daily drills on these. After all they are the foundations on which we target to erect the edifice of our learners' language proficiency.
3. University exam papers should be set on the same pattern as those of other international proficiency tests such as, MET, IELTS and TOEFL.
4. Recommendation 3 means that we have to upgrade our curriculum to prepare the learners for these. Here are some ways:
  - i. More comprehensive vocabulary lists for memorisation (as it is important to ensure a certain word bank in the learners' memory);
  - ii. Rigorous grammar training with planned increase in difficulty level;
  - iii. Writing practice on daily basis with greater emphasis on collaborative writing

Arabic is one of the official languages of the United Nations which offers fantastic employment opportunities to our Translation Graduates who may be employed as Translators and Precis Writers. [www.careers.un.org](http://www.careers.un.org) has posted sample examinations and previous years' question papers expressly for the purpose of practice for aspirants to these jobs.

1. Classroom modules must find space and time for these and other competitive exam practices. After all the idea behind all this language training is employability of our learners.
2. Speaking as a skill should find primacy in classes: three moderate learners mediated by one proficient learner should be encouraged to engage in speaking drills. Short group sessions should form part of the teaching schedule as a group of more than five EFL learners has been found contrary to conducive foreign language learning environment. Further, recent findings have shown that engaging three instead of the conventional two learners in speaking exercises is found to be more useful by the learners.
3. Assessments should be carried out at two levels: Formative, to evaluate if the learners are making expected progress in the specific language items being taught in the classroom; and Summative, to evaluate student learning and skill acquisition periodically.
4. Speaking is a skill that needs speaking test and not a written pronunciation drill as is done at KSA. During assessment, learners' proficiency for dialogue as well as reading aloud ability (to test pronunciation) should be tested.
5. In the benefit of the learners we must not shy from increasing the vastness and scope of the syllabus. Tutorials may be resorted to for better and individualised attention to each learner. A little extra effort in the present on the part of the teachers as well as learners can go a long way in ensuring a bright future.
6. Testing the Test Papers using the following Checklists:

##### Checklist 1. Face Validity of Question Paper

No.	Items	Yes	No	Suggestion(s) for Improvement, where applicable
1	University Name and logo			
2	Department name			
3	College name and location			
4	Semester No.			
5	Academic year			
6	Course name and code			
7	Course division			
8	Exam date			
9	Maximum marks			
10	Allotment of Marks (question wise)			
11	Duration of the exam			
12	Student's name & Academic no.			
13	Instructor's name			
14	Pagination			

##### Checklist 2. Qualitative Analysis of Question Paper (ideal responses are enclosed in brackets):

- i. Does the test rely heavily on the learners' expressive ability in English. (No)
- ii. Being Arabic speakers, are the learners familiar with a similar testing pattern in Arabic. (Yes)

- iii. Is English language proficiency a pre requisite for taking this test. (No)
- iv. Does the test address the specific item desired to be tested. (Yes)
- v. Are the questions asked familiar to the learners. (Yes)
- vi. Given that a foreign language is being tested, are there any confusing or unfamiliar contents in the paper. (No)
- vii. Were the learners exposed in the class to the items (both content and task design wise) (Yes)
- viii. Does the instrument test the learners' critical thinking ability. (Yes)
- ix. Does the instrument employ a variety of questioning procedures. (Yes)
- x. Is the test format familiar to the learners. (Yes)
- xi. On the counts of validity and reliability, is the test within acceptable limits for this particular learner group. (Yes)
- xii. Is the English vocabulary used in the test unfamiliar to the learners. (No)

### **Checklist 3. Analysing the Question Paper for Variety of Question Types**

**Kindly state whether the question-paper contains the following type(s) of questions:**

- i. MCQs
- ii. True/ False
- iii. Matching items
- iv. Fill-in-the blank
- v. Short descriptive questions
- vi. Essay type questions

### Ensuring Validity and Reliability of the Language Tests

Phelan and Wren define Reliability and Validity in 'Exploring Reliability in Academic Assessment' (2005-2006). A test is said to be reliable if the results it produces are consistent over a period of time. Test-Retest Reliability can be obtained by administering the same test twice over a brief time span. Parallel Forms Reliability is obtained by administering two versions of the same test. Inter-Rater Reliability measure the degree of agreement of assessment decisions of different raters. Finally, Internal Consistency Reliability measures the degree to which different test items produce similar results.

Validity tests the extent to which a test measures what it intends to measure. Face Validity as the name suggests ascertains the Validity (ie whether it covers all the items intended to be tested) simply by looking at the test. Construct Validity ensures that the measure actually measures the construct and not other variables. A panel of experts familiar with the construct can assess this type of validity. Criterion-Related Validity predict performance. Finally Formative Validity assesses the outcomes to surmise how well a measure is able to provide information to help improve the program under study.

### Limitations of the Study

The scope of the current study, however, is limited to evaluating the EFL assessment practices at Qassim University, KSA. The sample under consideration was only of eighty EFL learners. Though it can be assumed to be representative of the larger EFL learner community in the KSA, pedagogical and policy changes being serious issues need a greater number of studies.

## APPENDIX. FEEDBACK QUESTIONNAIRE

Here are some questions to obtain your views on the EFL tests that you take at Qassim University. You need to circle either Y (for Yes) or N (for No)

1. I understand the questions presented in the test. Y/N
2. The time allowed for the tests is sufficient for me. Y/N
3. I am aware that for employment I may have to take one of the international Proficiency Tests for English. Y/N
4. Testing time is one of anxiety and worry for me. Y/N
5. I think I do well on the tests. Y/N
6. I found the MET sample test rather easy. Y/N
7. The questions asked in the tests are similar to exercises I do in the class. Y/N
8. My performance would have been better had the test components been different. Y/N
9. I feel translation is an easy task for me. Y/N
10. For my aspirations from the EFL course, I find the tests right and closely related to real life situations I might use the language in. Y/N
11. The test contents were familiar to me. Y/N
12. I was able to tackle the MET sample test questions because I had studied these components in my course. Y/N
13. My teacher did not teach me in the class the way questions are asked in the MET sample test paper. Y/N
14. I feel the MET questions employ the language as I would be needed to in real life. Y/N
15. With my knowledge of English, I am confident of scoring well on any proficiency test for English. Y/N
16. Does the exam cover the major components of the syllabus?
17. Is there a distinction between the rubric and the question items with regard to the font size used?
18. Do the questions address the three main levels of students: above -average, average and below -average?
19. Are the questions context-based? (skill courses only)

20. Are the questions sequenced from easy to difficult?  
21. Is the paper marginalized?

## REFERENCES

- [1] Alderson, Charles J. (1990). Testing Reading Comprehension Skills (part two) *Reading in a Foreign Language*.
- [2] Alderson, J Charles, Grant Henning, and Yasmeen Lukmani. (1987). Levels of Understanding in Reading Comprehension Tests. Paper presented at Ninth Annual Language Testing Colloquium, Miami.
- [3] Alderson, J. C. & Urquhart, A. H. (1985). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing*, 2(2), 192-204.
- [4] Bachman, L. F., & Palmer, A. S. (1981). A multitrait-multimethod investigation into the construct validity of six tests of speaking and reading. In A. S. Palmer, P. J. M. Groot, & G. A. Trosper (Eds.), *The construct validation of tests of communicative competence*, (pp. 149-165). Washington, DC: TESOL Publications.
- [5] Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16, 449-465.
- [6] Bachman, Lyle F. (1991) Teachers of English to Speakers of Other Languages, Inc. (TESOL) What Does Language Testing Have to Offer? *TESOL Quarterly*, Vol. 25, No. 4 (Winter), pp. 671-704.
- [7] Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second-language teaching and testing. *Applied Linguistics*, 1 (1), 1-47.
- [8] Canale, M. (1983). On some dimensions of language proficiency. In J. W. Oller, Jr. (ed.), *Issues in language testing research* (pp. 333-342). Rowley, MA: Newbury House.
- [9] Cumming, A. (2009) What Needs to Be Developed to Facilitate Classroom-Based Assessment? *TESOL Quarterly*, 43(3), 515-519. Retrieved from <http://www.jstor.org/stable/27785032> Accessed May 7, 2017.
- [10] Erickson, M., & Molloy, J. (1983) ESP test development for engineering. MA: Newbury House.
- [11] Phelan, Colin and Wren, Julie. (2005-2006). Exploring Reliability in Academic Assessment [www.uni.edu](http://www.uni.edu) accessed December 6, 2016.
- [12] Rosenbaum, Eric. (1971). Foreign Language Testing *Die Unterrichtspraxis / Teaching German*, Vol. 4, No. 1 (Spring), pp. 1-6e.
- [13] Rossi, Gary J. (1983). Foreign Language Proficiency Testing: A Selected, Annotated Guide to Standardized Instruments, Rating Scales, and Literature *The French Review*, Vol. 56, No. 6 (May), pp. 840-858.
- [14] So, Youngsoo, et al. (2015). Using Trialogues to Measure English Language Skills. *Journal of Educational Technology and Society* Vol. 18, No. 2(April), pp. 21-32.
- [15] Troudi, S., coombe, C., and Al-Hamly, M. (2009). EFL Teachers' Views of English Language Assessment in Higher Education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546-555. Retrieved from <http://www.jstor.org/stable/27785037> Accessed May 7, 2017.
- [16] Weir, C.J. (1983). Identifying the language needs of overseas students in tertiary education in the United Kingdom. Unpublished PhD thesis, Institute of Education, University of London.
- [17] Xu, Y., & Liu, Y. (2009). Teacher Assessment Knowledge and Practice: A Narrative Inquiry of a Chinese College EFL Teacher's Experience. *TESOL Quarterly*, 43(3), 493-513. Retrieved from <http://www.jstor.org/stable/27785031> Accessed May 7, 2017.



**Fahad Saleh Suleiman Alfallaj** is an Associate Professor of English at Qassim University, KSA. He chaired the Department of English and Translation at the College of Science and Arts, Methnab, Qassim University, KSA for 3 years (2012-2015). And since then he has been involved in research work. He has published widely in International Indexed Journals. His research interests include Islamic translation, ESP, Learning Strategies and Theoretical Linguistics.



**Arif Ahmed Mohammed Hassan Al-Ahdal** is an Associate Professor of English at Hodeidah University, Yemen, and Qassim University, KSA. Though his vocation is teaching, Dr. Al-Ahdal is an avid poet and writer. He has to his credit two bi-lingual poetry collections, *To Yemen with Love and Joys and Sorrows*, published in Yemen in 2010 and a reference book for teachers "PREPARING EFL TEACHERS FOR TOMORROW: OPENING NEW VISTAS IN TEACHER DEVELOPMENT WITH INSET", published in Germany in 2015. He was born in Yemen where he developed a deep felt love for English which led him across nations to India. He pursued MA in TESL from the English and Foreign Languages University, Hyderabad, India, in 2008 and thereafter PhD in English from Panjab University, Chandigarh, in 2011. He also gained cross-cultural teaching experience from three countries: Yemen, India and Saudi Arabia. With

such diverse exposure, he was offered prestigious editorial responsibilities as Board Member of *ELT Journal*, Canada, *International Journal of Applied Linguistics and English Literature*, Australia, *Arab World English Journal*, Malaysia and *Journal of ELT and Applied Linguistics*, India, among others. He is also a freelance Trainer and Certified IELTS Examiner.