

An Investigation into the Comparability of English Language Exit Tests in a Number of Tertiary Level Institutions in Saudi Arabia

Ghurmallah Abdullah Alghamdi
King Faisal Air Academy

Abstract: *This paper attempted to examine the English language testing situation in some tertiary level institutions that use English as a medium of instruction in Saudi Arabia. The main question proposed is: to what extent do institutions in Saudi Arabia which are using English as a medium of instruction conform to the norms and standard of language tests' practices when designing language exit tests? For this purpose two questionnaires were designed. This paper concluded that reforms in language testing within some of the institutions studied are needed. These reforms should raise awareness among administrators and language testers to reconsider language testing procedures and testing qualities.*

1. Introduction

A number of institutions in the Saudi Arabia, such as King Fahad University of Petroleum and Minerals [KFUPM], King Faisal Air Academy [KFAA], the Institute of Public Administration [IPA], the Technical Studies Institute in Dhahran [TSI], Imam Mohammad University [IMMU] as well as the colleges of medicine kingdom wide are English medium institutions. We will take two institutions, King Saud University [KSU] and King Abdulaziz University [KAU] as examples to represent medical colleges. If tests used by these institutions do not reinforce teaching and learning by testing the language skills and language elements that students need, then there is a problem with content validity. If, on the other hand, relevant testing characteristics such as reliability, inter-rater reliability for subjective tests, (e.g. speaking and writing) and validity of tests are not established, then this poses an ethical problem. More importantly, if these tests are used without being subjected to the various testing designs, then the problem is even more critical. It is the aim of this research to find out the present state of language testing in the above mentioned institutions.

2. Study Objectives

The study aims at finding out the current state of English language testing in a number of English medium institutions in Saudi Arabia. It specifically seeks to find answers to a set of questions which relate to the preparation, administration and scoring of such tests. Study questions are listed in Section 5 below. For this purpose, the researcher developed two questionnaires (see Appendix II and Appendix III).

3. Context of Study

3.1. Institutions

All of the above mentioned institutions teach some specialties such as, medicine, flying training and technical training through the medium of English. They all provide English language training and require certain standards of language ability for students to pursue further training/education in English. Below is a brief description of the measures taken by each university/institute to upgrade students' proficiency in English.

- (i) KFUPM provides students with an orientation English course. At the end of this phase students are given qualifying tests. Those who pass these tests are eligible to study at one of the engineering colleges at the university.
- (ii) IPA gives students who will be studying "hospital management" a special course in English and requires a certain standard in English before they are admitted into this programme.
- (iii) The Colleges of Medicine and Applied Medical Sciences at KSU and KAAU require students to take and pass an English language test before they are allowed to proceed to the next phase of education.
- (iv) KFAA provides students with an intensive English language programme and requires them to pass an English language proficiency test before they proceed to the next phase of pilot training.
- (v) IMMU provides candidates with an English language course and at the end they are given a test before they are admitted into the English Department.
- (vi) TSI provides students with an English language course and requires them to pass a test before they are admitted to the second phase of Technical Studies.

As can be seen, these institutions have two things in common. First, they provide students with special English language programmes to prepare them for academic study/training. Second, upon completing this course they should pass a qualifying test before admission into academic programmes/training.

3.1. Characteristics of test-takers

Test takers in these institutions have many things in common: (i) They are all graduates of secondary schools, 18 to 19 years old, (ii) All are males except for the College of Medicine and Applied Medical studies at KSU and KAAU where girls are also enrolled in the English language programme, (iii) They all read for the BA or the BSc degree except for IPA and TSI students who read for a diploma, (iv) In total they had studied English as a compulsory subject at school for six years before joining their respective universities or institutes, and (v) All have to pass an exit achievement or proficiency test to qualify for the second phase in their university education : academic programmes in case of universities, and training in case of institutions.¹

Like other tests, language tests are important components of the learning-teaching situation. Hence, it is crucial that achievement/proficiency

language tests given as exit tests should conform to the standards set out by testing and measurement specialists, keeping in mind the needs of students and requirements of subsequent academic education/training. The current study aims at finding out to the extent these tests conform to the known standards in the field of language testing.

4. Literature Review

Literature review in this study is divided into two parts: (i) comparative studies and principles and (ii) studies on language test design.

4.1 Comparative studies

A major comparative study was conducted by Bachman, et. al (1995) to compare the Cambridge First Certificate in English (FCE) and the Test of English as a Foreign Language (TOEFL). The study concluded that, although the two tests were designed to test many of the same language abilities, they, nonetheless, exhibit significant differences in their approach to test design, reflecting major differences of approaches to language measurement between the USA and the UK. The following summarizes the most important aspects of their study:

- a. The qualitative content analysis of the two tests including the specific language abilities and type of test tasks,
- b. The characteristics of test takers, and
- c. The construct validity of the two tests.

The researchers concluded that they needed to place the comparison of the two language tests in the broader context of the philosophical differences between the British and American educational measurement systems. The study also pointed to several implications related to test users in the sense that this study revealed detailed information about the tests, a lot more than the ones published by the designers of these tests, which make it possible for administrators to base their decision in choosing a particular test on the basis of sound grounds. The study employed a variety of empirical approaches to research which, according to the researchers, has increased our understanding of the complexities of multiple datasets and how they are integrated as to allow for common analysis. For more details on implications for research, test users and for the field of language testing see Bachman, et. al., (1995:126-37).

According to Alderson; Clapham and Wall (1996:259) principles and standards by which tests can be compared still require consensus among language testers. They, for instance, note that "Language testing lacks any agreed standards by which language tests can be evaluated compared or selected." (ibid: 259). However, it is our contention that principles of comparison can be derived from the qualities of a good language test as well as from differences in test content in so far as abilities, elements and skills being tested are concerned. These principles may include: (a) abilities being tested, (b) validity, (c) reliability, (d) practicality and feasibility, and (e) wash-back.

4.2 Test design

Before one embarks on designing an exit test, he has to work within the Critical Thinking approach (cf. Shohamy, 2001). This approach, according to Shohamy (1995) and Pennycook (1994) proposes the following directions (see Shohamy 2001, 131-2):

1. The act of language testing is not impartial. It is a product that reflects societies' educational values. It also has an effect on the lives of test takers, teachers and educators. And, these are tools related to levels of success.
2. Test takers should be encouraged to develop a critical view of tests.
3. Traditional Psychometric theory of test design does not represent the absolute truth.
4. Language testers' knowledge is incomplete; therefore we should also seek knowledge from additional sources in order to get more valid interpretation of data.
5. Results obtained from research on language testing should be open to discussion with many parties.

In Davies' (1997:328) words "Critical approaches to language testing expose the importance of carefully examining alternative assessment proposals and making clear the validity of the assessment methods used by the profession." Davies suggests that a language test should not only prove to be valid and reliable but this validity and reliability should be explained in a way that leaves no doubt about the design of the test by asking rigorous questions similar to those asked by Shohamy (2001). Generally speaking, her questions specify clearly all steps involved in language testing, not only the practical steps in the process of language test design, but those that could be addressed by students and other test users. The following are examples of these questions:

Who are the testers? What is their agenda? What is not being tested and why? What are the underlying values of the test? Can the test, its results and rationale be challenged? What ideology is delivered through the test?

We can also add other questions that relate to ethical considerations such as: To what extent does a language tester conform to the ethical code published by the International Association of Language Testers.

The steps of test design can be summarized as follows:

- (i) Test specifications should be considered in view of the following questions: What is the purpose of the test? Who are the test takers? What language skills and elements should be tested? What are the test methods? What criteria will be used for assessment? What is the items format?
- (ii) Other relevant processes involve: retesting and analysis of items, which involves preliminary trials of the test, major trials, item analysis and descriptive statistics, reporting scores and setting cut offs and finally the validation processes.

Many language testers ignore an important stage of test design, which is examining the wash-back effect of the test. This stage is as important as the other stages referred to above. Alderson, et. al. (1996) provide an excellent and detailed review of the various stages of test design and suggest ways of comparing tests. However, it is beyond the scope of this paper, to go into such details.

5. Study Questionnaires

5.1. Questionnaire design

The questionnaire used in this study aimed to establish the qualities of the English language tests used by each of the institutions involved in this study for the purpose of comparing the different tests. It consists of 25 items which seek answers for the following general questions (See Appendix (I) for details):

- What type of test is being used in each institution?
- Has it been locally designed or adapted?
- Is it computer based or a paper/pencil type test?
- How long does it take to complete?
- What language skills and/or elements are tested in each institution?
- What format(s) does each test exhibit?
- What is the number of items in each subtest and the test as a whole?
- What are the scoring procedures used?
- Has the test been validated?
- Has its reliability been established?

The questionnaire was first checked for content and clarity of language by a native speaker of English who has a PhD in linguistics and who is also responsible for testing in KFAA.

5.2. Questionnaire administration

Two methods of distribution were used. First, questionnaires were either handed in person or sent by electronic mail [to the head of English Department in each institution] and was accompanied by a cover letter illustrating the purpose of the research project. The second part of the questionnaire was (wherever applicable) completed by the language tester in each of these institutions. Each institution completed and returned the questionnaire within a fortnight, except for TSI, which took longer.²

6. Data Analysis and Discussion

Appendix (1) below gives the response of all institutions to the study questionnaire with regard to their exit tests. However, before we discuss the testing situation in each institution separately, two observations need to be made:

- (a) It seems that each institution is satisfied with the exit tests they are using.
- (b) Some institutions have not provided the evidence on which they decided that their exit tests were valid and reliable.

7. Results and Discussion

Below is a brief presentation of the responses of the various intuitions at which the study questionnaires were administered.

7.1. Types of tests: Three of the institutions label their exit tests as "achievement test" and the other three label their exit tests as "proficiency test". See Appendix (I) for details.

7.2. Tests titles: All institutions have titles for their tests, except for one. (See Appendix I below).

7.3. Tests source: All institutions contend that their tests are locally designed. This, in our opinion, should be the case for obvious reasons. Adapted tests might not be suitable for cultural reasons but most importantly content wise, in so far as skills and language elements tested are concerned.

7.4. Team vs. individual work: The literature of language testing stresses the fact that no single person can construct a language test without some help from other colleagues at least in the form of review. Reviewers as outsiders might be in a position to observe what the item writer takes for granted. They also look at test items from the point of view of the test takers not the test designers. In this respect Alderson et al (1996) note that "As we have repeatedly emphasized, no one person can possibly produce a good test, or even a good item, without advice" (62). It is, therefore, of crucial importance that test construction, especially those that are going to decide the future of test takers, must be reviewed by other specialized persons immediately after the test items are prepared and before piloting the test. This stage is known as the editing stage. It is reassuring that all exit tests in the institutions covered in this study are the result of team work.

7.5. Number of versions: All institutions, except for two (see Appendix (I) below) have more than one version. TSI, however, has the largest number of versions (30 in total). Experience as well as logic suggest that the fewer the number the versions, the more likely they will lend themselves to overuse, lose reliability, which will eventually lead the test to be compromised and thus lose reliability.

7.6. Methods of testing: Some institutions made use of technology, their tests are fully computerized. However, other institutions still use "paper and pencil" tests. Computers are particularly useful for large scale testing. TSI, for example, has about three thousand students to test every year. Therefore their use of computers saves them a lot of time. Computers are also very useful when making test analysis.

7.7. Length of tests: KAAU and KSU tests take three hours to complete; tests in the other institutions take between one to 1.5 hours. The length of a test depends on the number of items as well as the number of elements and skills tested. However, it is established in the literature on language testing that the longer the test, the more reliable it becomes.

7.8. Skills tested: It seems that every institution recognizes very well what the specific skills which their students need to cope with their future study. The skill of speaking is particularly important for pilots, for example, therefore KFAA requires its students to pass a speaking test. On the other hand, reading is very important for the students joining medical colleges because they will be required to read and process many references about their specialty through the medium of English. Almost all skills are tested by every institution.

7.9. Tests format: It seems that the most popular testing format in the institutions of this study is the multiple-choice test. This technique has some advantages but also some disadvantages. Ease and objectivity of scoring is one advantage; another is that it can be used for large scale testing. On the other hand, it is very difficult to construct and when it is used to test language, it does not tell us whether the candidate can use the language for communicative purposes.

7.10. Number of items: The total number of items varies from one institution to another. The TSI test has the largest number of items (200), whereas the KFAA test has the least, (viz 75). On the other hand, the KAAU and IMMU tests have equivalent number of items (90), whereas the IPA and KSU tests have 118 and 108 items respectively.

7.11. Speaking tests: Only two institutions administer speaking tests, namely IPA and KFAA. Other institutions do not require their students to sit for speaking tests, perhaps not because their students do not need oral skills but may be because it is impractical for them due to certain educational constraints such as time constraints or lack of sufficient numbers of examiners. The IPA test lasts for five minutes [including re-telling of a story or talking about a series of pictures], whereas the KFAA speaking test takes the form of an oral interview which lasts for 10-15 minutes.

7.12. Writing tests: Three of the institutions give writing tests: IPA, KAAU and KSU. IPA requires their student to write a letter of no less than 200 words. KAAU, on the other hand, uses the technique of information transfer, such as diagrams where students are required to transfer information into a written text in no less than 100 words, whereas IMMU requires candidates to write 10 sentences about certain expressions as is the case in the TOEFL.

7.13. Methods of scoring MCQS: Four institutions use machines to score mcqs and two score them manually. However, the two methods require some kind of

checking procedure by which one makes sure that results are accurate before they are announced.

7.14. Speaking and writing tests: Not all institutions participating in this study test their students in writing or speaking. The marking of these tests is highly subjective and requires some kind of scales by which assessors can make judgments about candidates' performance. Two types of scales have been developed within the literature of language testing, namely the analytical and the holistic scales. An analytical scale assesses students' performance in terms of separate linguistic elements such as grammatical accuracy, pronunciation, range and appropriate use of vocabulary, etc, whereas the holistic scale assesses students' performance in terms of global descriptors, where performance is assessed against an overall criteria which includes all features of possible performance. Four institutions in this context use the holistic approach. One institution (KAAU), however, uses double-marking for scoring writing, the final mark being the average of the two marks.

7.15. Establishing reliability: Only two institutions provide information about the reliability of their tests: IPA and KFAA. IPA used parallel versions to establish reliability of the test, while KFAA used KR-21 to establish the internal consistency of their tests. Other institutions have not worked out the reliability of their tests.

7.16. Content validity: All institutions claim that their tests are content valid. Verifying these claims is beyond the scope of this paper. However, for a test to have content validity, it has to appropriately reflect and sample a certain defined syllabus.

7.17. Sampling domain: IMMU claims that its test is a proficiency test; therefore it samples the whole domain of the language. Other institutions contend that their tests are achievement tests and reflect a certain syllabus. KFAA, on the other hand, say that their test is a proficiency test although it reflects a defined syllabus.

7.18. Predictive and concurrent validity: None of the institutions validated its test in terms of predictive or concurrent validity. Here, we are talking about exit tests that decide candidates' future, therefore beside content and construct validation, predictive validity needs to be established, otherwise, we are not sure that we are making the right judgment.

7.19. Construct validation: Only two of the institutions worked out construct validity: KFAA and IPA. For this purpose, two methods were used; factor analysis and sub-tests intercorrelations. Factor analysis revealed one general factor and the sub-tests intercorrelations suggested that each test correlated

moderately with another indicating that they share some linguistic characteristics with each other.

7.20. Statistical analyses: It seems that most of the institutions did not do much statistical analyses. Even item analysis was not done, let alone calculating reliability and working out validation procedures. When asked how they decide on what is considered a bad item, they said: when item reviewers feel that an item is not good, we do not use it. In other words they rely on intuition.

8. Follow-up Questionnaire

A follow-up questionnaire was designed and sent to the same institutions which responded to the first one. This time the questionnaire had eight items. It was aimed at finding out more about the actual steps undertaken during test construction. As with the previous questionnaire, this one was also checked for content and language by the same professor, no amendments were made on language, but this time a couple of questions were added. (See Appendix (III) for content). Results will be presented and discussed below.

8.1. Needs analysis: Needs analysis is a basic requirement prior to designing tests, especially those that will decide the future of the candidates or the direction of their study. It is true that needs analysis is done mainly prior to the construction of test materials; however it forms the basis for language test analysis. When asked whether any form of needs analysis had been conducted to help decide upon the purpose, content, method and level of the test, responses were as follows: (i) KAAU, IMMU and IPA responded that no form of needs analysis had been conducted prior to the test design; (ii) KFAA and TSI said that needs analysis was done prior to the construction of their test.

8.2. External tests vs. local: KAAU indicated that they did not consider using an external test because they thought it would be more appropriate to design a test based on their curriculum for Medics. KFAA, on the other hand, said that the only test they were in a position to consider was a new version of the American Language Placemen Test [ALCPT], which they believed to be statistically reliable but had an unnecessarily heavy vocabulary load, and did not test grammar, listening, speaking and writing. IPA said that they did not consider using an external test because the test they use is dependent on the program they offer, whereas IMMU said that they considered using an external test, but gave no further explanation and TSI indicated that the only test they considered using was the ALCPT but, because it was course-bound, they thought it was unsuitable.

8.3. Theoretical framework: Apart from KFAA, all other institutions said that their test reflected no specific theory of language or language learning. However, KFAA said that the basic design of their test reflected the notion of a core

vocabulary of very frequent lexical items derived from the work of the COBUILD group. Generally speaking, however the listening component owed something to Relevance Theory. (cf. Wilson and Sperber, 1993).

8.4. Test and curriculum: All institutions asserted that their tests are very closely related to the curriculum. Vocabulary and grammar were stored in a data base, and their position in the course logged. With regard to skill testing, the test formats were developed in conjunction with course developers. IMMU said that the test was based on the textbooks of students. Similarly, IPA noted that the test items are taken from the exercises in the student textbooks and KAAU said that the test directly relates to the course.

8.5. Guidance to test writers: Below is a summary of the responses of the various institutions with regard to providing guidance to test makers.

(i) KAAU said that writers were given a list of core vocabulary items, and could not use any other lexical items. They were also supplied with definitions and example sentences taken from the COBUILD dictionary. For the grammar test, they were given descriptions of grammar areas to be tested. For the listening test, they were either given scripts or asked to develop scripts or unscripted dialogues according to varying precise guidelines. Also, test writers are chosen because they are valued as classroom instructors or because they have been working in particular skills in course development. Two of the multiple choice writers were native speakers of Arabic and this was one criterion used in their selection.

(ii) IPA respondent said that test writers were not given any guidelines. Also test writers are assigned according to the skill they teach.

(iii) KAAU indicated that they asked test writers to carefully examine the curriculum and the exam was checked for its direct link to the course and all class teachers contributed to sections of the exam.

(iv) TSI indicated that most test writers were native speakers with many years of test writing experience. They gave advice to non-native speakers and further guidance was given relevant to computer-based testing as well as test writing.

(v) IMMU said that no guidance was given to test writers and noted that they only have one lecturer who is a language tester.

8.6. What happens once a test draft is produced?

(i) KAAU said that it is normally reviewed first by a panel then pretested.

(ii) TSI said that tests are written by one team and then reviewed by another team. Following any amendments undertaken at this stage, the tests are pretested. Subsequently there may be one or two re-writes of some of the questions with further trailing.

(iii) KFAA said that the draft is first reviewed and edited by a panel then pretested in this case the panel consisted of the senior instructor responsible for testing. IMMU and IPA said that a draft is normally reviewed and edited by a panel then pretested.

8.7. Pretesting procedures:

- (i) IMMU indicated that their test was not trailed at all, i.e. given to a representative sample as part of the validation process to check on the validity of the items and of the test as a whole.
- (ii) TSI and KFAA indicated that their test was trailed on the students who represent the intended level.
- (iii) KFAA respondent said that in so far as this was possible they piloted the test on the entire intake when they reached the intended level.
- (iv) TSI trials involved between 150 to 200 cadets.
- (v) KFAA said that the number of students taking the pretest varied between 100 and 150 students.
- (vi) IMMU said that no procedures of pre- and post-testing were done.
- (vii) IPA indicated that their test was trailed on the intended level and the number involved 25 students for each trial.
- (viii) KAAU indicated that their test was not trailed.

9. Conclusion and Recommendations

This study indicated that English language tests in some institutions mapped in this study do not conform to the basic principles and procedures of language testing development. Many of the institutions are using tests that are not validated to decide if candidates are likely to pass in a programme taught through the medium of English. Even traditional item analyses were not done for some tests. Therefore, we cannot be absolutely certain if these tests sample the domain of the language properly and how valid and reliable they are. However, all institutions contend that their tests have content validity and what is being taught. This is particularly noticeable from answers to the follow up questionnaire.³

The main findings/ recommendations of this study are:

1. Some tests do not conform to the principles of basic test design. It is imperative that the basic principles of foreign language testing be faithfully adhered to.
2. Tests that decide candidates' future should be validated before they are put to use.
3. Some institutions have only one or two versions of the test which with overuse will be compromised. Depending on how frequent these tests are used, sufficient versions should be designed.
4. Further research is needed to establish content and construct validity of the tests in question, which will necessarily lead to the examination of both design and procedures.
5. Test users need to pay special attention to the effect of the tests on the teaching/learning situation. Promoting positive effects of the tests on instruction and curricula is of paramount importance.

Finally, this paper may be seen at the outset as overemphasizing statistics at the expense of content and construct validation processes. To the contrary, we consider this the first step which hopefully will lead to further research to address other aspects of the matter. However, statistics and figures are vital components of test validation processes.

Notes

¹ Students of medicine at KSU take an English language programme for one semester, some students do not pass in the first semester, so they are given an additional semester and have another attempt at the test before they are allowed to proceed to the following phase of their study. KFAA students, depending on their English language background, study English for a period of 10-42 weeks. TSI students study English for a maximum of 65 weeks. The orientation programme at KFUPM allows students to study English for a maximum of two semesters. As for IMMU students, who will eventually join the English Department, they study English for one semester.

² According to KFUPM, they chose not to complete the questionnaire because they had not yet conducted extensive statistical analyses on their exit test, so they have been excluded from this study.

³ The study is limited only to proficiency and achievement tests used as exit tests in KFUPM, the Colleges of Medicine at King Saud and KAAU universities, IMMU, KFAA, the (IPA) and TSI.

References

- Alderson, C., Clapham, C. and Wall, D. (1996).** *Language Test Construction and Evaluation*. Cambridge University Press.
- Bachman, Lyle .F (1990).** *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L., Davidson, F., Ryan .K and Choi, I. (1995).** *Studies in Language Testing 1: An investigation into the comparability of two tests of English as a foreign language*. Cambridge University press.
- Davies, A. (1997).** Demands of being professional in language testing. *Language Testing vol. 14, 3*. pp. 328-339.
- Pennycook, A. (1994).** *The Cultural Politics of English as an International Language*. London and New York :Longman.
- Shohamy, E. (2001).** *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. Pearson Education Limited.
- Wilson and Sperber, D. (1993).** Linguistic form and relevance, *Lingua*, 90, 1-25.

Appendix I: at the end of the paper**Appendix II: First Questionnaire**

1. Our institution is: KFUPM [] ; KSU [] ; IMMU [] ; IPA [] ; KFAA [] ; TSI [] .
2. The exit test we use at our institution is: achievement [] proficiency [] .
3. The title of the test we use is: -----
4. The test we use is : locally designed [] adopted or imported [] .
5. If locally designed was the test designed by: one Language tester [] // a team of language testers []?
6. The test we use exists in [] versions. Please specify.....
7. It is: a computer based test [] a paper and pencil test [] a mixture of both [] .
8. It takes: 1 hour [] ; 1.5 hours [] ; 2 hours [] ; 3 hours [] to complete.
9. It tests the following skills and components: listening [] speaking [] reading [] writing [] ; vocabulary [] ; grammar [] .
10. The skills being tested exist in the following formats:
 - Listening: short dialogues followed by multiple choice [].; visual related discourse followed by multiple choice [] ; a short lecture followed by multiple choice [] ; information transfer [] ; others, please specify:-----
 - Speaking: oral interview [] ; retelling of a story [] ; talking about a series of pictures [] ; others: please specify: -----
 - Reading: texts followed by questions [] ; cloze [] ; information transfer [] ; others: please specify [] .
 - Writing: an essay [] ; letter writing [].; others, please specify: -----
 - Vocabulary: multiple choice [] ; others, please specify: -----
 - Grammar: multiple choice [] ; others, please specify: -----
11. The number of items in each sub-test is: listening [] items.; reading [] items; grammar [] items; vocabulary [] items.
12. The speaking test lasts for [] minutes. Please specify.
13. The writing section requires students to write not less than: 100 words [] ; 200 words [] ; more than that [] ; a different number. Please specify []
14. Scoring of multiple choice items is done by: machine [] ; manually []
15. For speaking we use : An analytical scoring procedure [] ; Holistic [] . For writing we use : an analytical scoring procedure [] ; holistic []
16. Items were validated in terms of discrimination and facility indices: Yes [] No [] .
17. The reliability of the test has been established using one ore more of these procedures: 1. test re-test [].; 2 split-half [] ; 3. internal consistency [] ; 4. parallel versions [] .

18. The reliability of the test and sub tests is as follows. Please specify a statistical figure: the whole test []; listening []; grammar []; vocabulary [].
19. Inter- rater reliability of the following skills is as follows: Please specify figures. speaking []; writing [].
20. We believe that our test is valid. Yes [] No []
21. Our test was validated in terms of: content []; construct []; prediction []; concurrence [].
22. Our tests properly reflect: a particular language course []; the whole domain of the language [].
23. Pearson Product Moment Co-relation Coefficient was used to establish:
- predictive validity [] and the figure is []. Please specify.
 - concurrent validity [] and the figure is []. Please specify.
24. Construct validity was established using factor analysis [] or [sub-tests inter-correlation of the Pearson Product Moment [].
25. Factor analysis shows: one main factor only [] or multiple factors [].
26. Figures of the inter-correlations between sub-parts is as follows: grammar and vocabulary []; listening and speaking []; reading and writing [].

Appendix III: Second Questionnaire

OUR INSTITUTION IS -----.

1. Has any form of needs analysis been conducted to help decide upon the purpose, content, method level of the test?. Yes () No ().
2. Did you consider using any external test? If you decided against using an external test, please give reasons.
3. Does the test operate within the framework of any theory of language or language learning? Yes () No (). If yes please specify.
4. How was the test linked to the course curriculum?
5. Are test writers given any further information or guidance? Yes () No (). How are test writers chosen? Please specify the nature of information or guidance used in the selection.
6. Once the first draft has been produced, what then happens to it? Reviewed and edited first by a panel then pre tested () or pre tested straight away ().
7. How are students selected for the purpose of pre testing? Students who represent the intended level () any student in the institution regardless of his level () or not trailed at all ().
8. How many students were used in the trailing of the test?

Appendix I: Information on Exit Tests in KSA

	IPA(Hospitals administration)	KFAA/Flying training	KAA/Medicine	IMMU/English	KSU/Medicine	TSI/Technical training
Type of test	Achievement	Proficiency	Proficiency	Proficiency	Achievement	Achievement
Title of test	IPA English Language Test	Core Proficiency Test	Medics Final Examination 1 st year	NA	Medics Final Examination	English Language Comprehension Test
Origin	Locally designed	Locally designed	Locally designed	Locally designed	Locally designed	Locally designed
Designers	Team work	Team work	Team work	Team work	Team work	Team work
Number of versions	Two versions	Eight versions	One version	One version	Four versions	Thirty versions
Methods of testing	Paper & Pencil	Computer based	Paper & Pencil	Computer based	Paper & Pencil	Computer based
Length	1 hour	1.5 hours	3 hours	1.5 hours	3 hours	1 hour
Skills tested	Listening, Speaking, Reading, Grammar & Vocabulary	Listening, Speaking, Vocabulary & Grammar.	Reading, Writing, Vocabulary & Grammar	Reading, Writing, Vocabulary & Grammar	Reading, Listening, Writing, Grammar & Medical Terminologies	Listening, Reading, Vocabulary & Grammar
Format	Listening: short dialogues followed by MCQS & short lecture by MCQS. Speaking: Retelling of a story & talking about a series of pictures. Reading: Texts followed by questions. Writing: letter writing Grammar & Vocabulary: MCQS.	Listening: short dialogues followed by MCQS, visual related discourse followed by MCQS & a short lecture followed by MCQS. Speaking: oral interview. Vocabulary: MCQS. Grammar: MCQS.	Reading: Texts followed by questions Writing: Information transfer, making sentences/ paragraph from diagrams & graphs. Vocabulary: MCQS Grammar: MCQS	Reading: Texts followed by questions Writing: expressions like TOEFL. Vocabulary: MCQS Grammar: MCQS	Reading: Texts followed by MCQS, True false & information transfer Medical terminology: Fill in blanks, MCQS & True/false Listening: Texts followed by MCQS & True/False. Students are allowed to take notes while listening. Grammar: MCQS & Verb correction Writing: Guided composition.	Listening: Short dialogues followed by MCQS Vocabulary: MCQS Grammar: MCQS

Number of items	Listening: 24 items Reading: 35 items Grammar: 35 items Vocabulary: 24 items	Listening: 25 items Grammar: 25 items Vocabulary: 25 items	Reading: 20 items Grammar: 25 items Vocabulary: 45 items	Reading: 30 items Grammar: 30 items Vocabulary: 30 items	Listening: 20 items Reading: 18 items Grammar: 50 items Terminology: 20 items	Listening: 50 items Reading: 50 items Grammar: 35 items Vocabulary: 65 items
Length of speaking test	5 minutes	10 to 15 minutes	NA	NA	NA	NA
Writing	200 words	NA	100 words	NA	10 sentences [less than 100 words.]	NA
Methods of scoring MCQS	Manually	Machine	Machine	Machine	Manually	Machine
Types of scales used to score subjective tests	Speaking: Holistic Writing: Holistic	Speaking: Holistic	Writing: Holistic	NA	Writing: Holistic	NA
Method of establishing reliability	Parallel versions	Internal Consistency KR-20	Have not been worked out	Have not been worked out	Have not been worked out	Have not been worked out
Reliability	The whole test: 0.80 Listening: 0.30 Grammar: 0.60	Vocabulary: 0.75 Grammar: 0.76	Have not been worked out	Have not been worked out	Have not been worked out	Have not been worked out
Inter-rater reliability	Speaking: 0.75 Writing: 0.75	Speaking: 0.81	Writing: double marked	NA	NA	NA
Validity	Content valid	Content valid	Content valid	Content Valid	Content valid	Content valid
The test reflects	A particular Syllabus	A particular Syllabus	A particular language syllabus	The hole domain of the language	A particular syllabus	A particular syllabus
Predictive & concurrent validity	Have not been worked out	Have not been worked out	Have not been worked out	Have not been worked out	Have not been worked out	Have not been worked out
Construct validity	Factor analysis: One factor	Factor analysis: One factor	Have not been worked out	Have not been worked out	Have not been worked out	Have not been worked out
Test Interco relations	Grammar & Vocabulary: 0.30 Listening & Speaking: 0.70 Reading & Writing: 0.70	Grammar & Vocabulary: 0.77	Have not been worked out	Have not been worked out	Have not been worked out	Have not been worked out