

The Entry Test of the English Department at the College of Arts: Evaluations and Adoption

Abdulmir Alamin¹ & Sawsan Ahmed¹

¹ Department of Foreign Languages, College of Arts, Taif University, Taif, KSA

Correspondence: Abdulmir Alamin, Foreign Languages Department, College of Arts, Taif University, Taif, 21974, KSA. E-mail: alamin@tu.edu.sa

Received: March 7, 2012

Accepted: March 23, 2012

Published: June 1, 2012

doi:10.5539/jel.v1n1p13

URL: <http://dx.doi.org/10.5539/jel.v1n1p13>

Abstract

The purpose of this paper is to provide guidance and criteria for the Department of Foreign Languages at Taif University, KSA, to be more selective in choosing students to study English language and literature. This study is also an evaluation of the courses offered by the department.

Furthermore, the objective of this study is to investigate the feasibility of using effective and reliable testing tools to assess and evaluate students' performance and accomplishment. This test will help to have a better idea about students' performance and competence. This will give the policy makers at the Department better understanding and knowledge for the process in selecting the courses and reference books as well as tailoring the teaching materials towards students' needs and their career. Last, this will help the people in the authority to pinpoint the weaknesses and strengths of the programs offered by the university.

Keywords: assessment, evaluation, criterion-referenced, facility index

1. Introduction

1.1 Aims of Research

The aims of the research are to achieve the following:

- 1) Explaining the difference between assessment and evaluation.
- 2) Explaining the difference between "formative evaluation" and "summative evaluation".
- 3) Outlining the different basic methods of carrying out an evaluation.
- 4) Pointing out the extent to which the Entry Test of the English Department can achieve the intended objectives set by the admission committee for setting up the test items.
- 5) Enabling the department to find out how effective and successful the Entry Test as a means of measurement to determine the best candidates to join the department to study English language.

1.2 Objectives of the Research

They are to:

- 1) Analyse students' answers to the questions of the Entry Test.
- 2) Recognise the "easiest" and the "most difficult" items of the above mentioned test.
- 3) Enhance decisions about how to create and improve more effective admission test.
- 4) Which part of the Entry Test is the most successful?
- 5) Which part is the least successful?
- 6) To point out the good and bad questions.

1.3 Limitation of the Research

The research was limited to assess and evaluate the Entry or admission test of the English Department, at the faculty of Arts, Taif University. This criterion-referenced test is usually given to students, who are willing to join

four- year BA program. The results of this Entry Test would be useful for selecting the best candidates to study at the department.

2. Evaluation and Assessment

2.1 Language Assessment

Many aspects of the language study may be tested at a number of levels and in a variety of ways. Most of teachers continue to set up tests of a kind they are familiar with without asking themselves these basic questions:

- 1) What is my purpose in testing these students?
- 2) How is the test related to the objectives of this course?
- 3) What do I expect this test to show?
- 4) Am I really testing what the students have been learning?

Answering these questions will involve a fundamental understanding of the principles of both assessment and evaluation! There is a useful distinction we can make between these two words which in every day and indeed in most dictionaries seem hypothetical. If we know exactly what it is synonyms that a student is expected to be able to do, then all we have to do is to ask him or her to demonstrate that he or she can do it in the forms of “tests” or “examination” which both means of assessment. Thus, ASSESSMENT, as in Oller (1978), Rivers (1968), and Rountree (2008), is the process of finding out what the students have attained to what level of proficiency.

2.2 Kinds of Language Assessment

Assessments vary in nature. If it is to rank or compare the achievement of student in relation to other students, this will be called NORM-REFERENCED ASSESSMENT. Norm referenced assessment, therefore, is competitive since students effectively compete with each other for relative placing in a rank order of merit.

Otherwise, if criteria, sets of objectives, which the students must each attain, are established and each student is expected to master all of the objectives, CRITERION-REFERENCED assessment is used in relation to the attainment of specified objectives. Assessment may be formal or informal. It may be conducted throughout a course on a continuous or intermittent basis, or it may take place at the end of a course or both.

The common forms of language assessment include the following:

1) Traditional Examination

It is well-known form of examinations in which students are asked to write essay-type answers within a limited time to a selected number of questions. These questions sample the subject content of the course and they are usually regarded as being easy to construct but difficult to mark and there is great scope for subjectivity in deciding the merit of the answers.

2) Oral Examination

Oral examinations are well known in language testing-objectives need to be precisely specified.

3) Objective Tests

The most frequently used question types include: multiple-choice, multiple-completion, matching pairs, and true/false. They are “objective” in two senses. On the one hand, they are associated with particular language course objectives and on the other hand, they require no subjective interpretation of the answer and no skill to mark. Objective test questions are usually called items because they need not necessarily be actually in the form of questions they must above all be such that the correctness of the “right” answer is unarguable and completely acceptable to a language specialist.

4) Comprehension Tests

Comprehension tests can take a variety of forms but the typical kind consists of a short reading passage of text followed by a number of questions which require the learner to show his or her comprehension of the text.

5) Close Tests

Usually a passage of text has words deleted and the learner is required to supply either the exact missing words, or words that are contextually appropriate.

2.3 Language Evaluation

Evaluation, on the other hand, is an attempt to identify and explain the effects (and effectiveness or appropriateness) of the teaching program or any education system. If the results of carrying out an evaluation are

to be any real value, then the procedures for obtaining those results must be beyond reproach. The whole process must be carried out systematically in such a way that the out-come is reliable and that the findings may be general.

When planning an evaluation scheme, it is important to consider the following:

- 1) Why the evaluation is being carried out?
- 2) For whom the evaluation is required?
- 3) What is to be evaluated?
- 4) When the valuation is to be carried out?
- 5) How the evaluation is to be carried out?
- 6) How information is information be collected an analysed?
- 7) What questions are to be asked?
- 8) How the questions are to be formulated?
- 9) Who will carry out the evaluation?

2.4 Kinds of Language Evaluation

Whether evaluation is Internal- that is carried out by that teacher who is implementing a new course, or a different teaching method, or who is trying out a new teaching, and so on. Or External--that is carried out by someone or some official body not involved in the design or implementation of the system being evaluated. It is of importance that the results should be as a reliable as it is possible to achieve.

The American literature has developed a distinction between two types of evaluation. Thus, "formative evaluation" is intended to develop and improve the "system" while it is still at a stage when changes can be made. An informal kind of formative evaluation is carried out every day by teachers who modify their teaching in response to the way their students are learning. If students fail to grasp what the teacher wishes them to learn, then teacher tries a different teaching strategy.

"Summative evaluating" enables judgements to be made about how well a system is or was in achieving the course objectives. Evaluation carried out for educational research is often summative, and this research is no exception.

The most common methods of carrying out evaluations are questionnaire and test or examination results. The way that questionnaire items are reworded is vitally important. The questions should not lead the students, or be emotionally biased in any way. On the other hand, pre-tests and post-tests can be used if the purpose is to measure learning gains. Otherwise, Facility Index and Discrimination Index are used if the purpose is to judge the effectiveness of a test. Facility index gives the proportion of student who obtained the right answer; the Discrimination Index is a measure of how well a test item can distinguish between the more able student and the less able.

3. The Entry Test of the English Department

3.1 The Entry Test Format

Partly because of the general poor standard of the secondary school leavers in the performance of English language, and as an antidote to the drawbacks of the current school teaching syllabuses, English Department at the Faculty of Arts would set up a policy that students would not join the department unless they prove their eligibility and efficiency in English Language.

Henceforth, an entry test format was built by certain staff members to cover reading, writing, speaking and listening skills. Grammar was included too. The format of the Entry Test under investigation consists of three sections: Section one is reading comprehension. This was set up into two parts: part one was a reading passage followed by ten wh-questions. Part two was multiple-choice questions. It aimed at testing students' ability in reading. Thirty marks were allocated for these two parts. Section two is grammar. This section covered most predetermined student's knowledge of preparatory and secondary school level. It aimed at ascertaining a student's prerequisite of English structure. It was designed into eight filling-up space sentences. They were about; present simple tense, present continuous tense, future tense, present perfect tense, past simple tense, past continuous tense, past perfect tense, passive voice, direct speech, mass and unit, how many, how much, conditional sentences and prepositions. Fifty marks were allocated for this section. Section three is writing. In order to assess students' ability in writing a proper English, they were asked first to write a short letter inviting a

friend, secondly, to write a short paragraph about on topic, a choice was given here to choose one out of three. Ten marks were allocated for each. Since this question was very subjective. Thus, there should have been some kind of agreement on basic ideas if model answers were not possible. Unfortunately none of these were available.

3.2 Validity

When speaking of validity in foreign language tests, it is important to ask: “valid for doing what?” This brings us firmly back to the need to be quite clear about objectives. With the existing English Entry Test, its objective is quite clear, that is simple for selecting students with an acceptable standard of English. This policy has been adopted by the Department of Foreign languages at Faculty, hopping to contribute in tackling teaching and learning problems of English in the country.

The English Entry Test has a content validity- since it was written by six teaching staff members, all of them are experienced and specialists in the field. They agreed collectively upon the items. They can be taken as judges for finalising and accepting the final form of the test.

3.3 Reliability

The 91 answered sheets were divided into odd numbered items and even ones. The mean and the standard deviation for each test were calculated between the two forms of tests. It was 0.70. Then the Spearman- Brown formula (Rowley, 2005) was used to get reliability 0.82. It is noticed that the reliability is high but it is still less than one, due possibly to the considerable proportion of unrecompensed items. Only twenty students out of the ninety one, whose marks in the second test, were more than that of the first one. Otherwise the majority answer swung toward the first test. (See table 1).

Table 1. The reliability of the admission test

Tests	No	Mean	Standard Deviation	Correlation between the Tests	Reliability
The First Test		12.49	5.63		
	91			0.70	0.82
The Second Test		10.83	5.41		

3.4 Time

The time for answering the written examination was decided to be two hours and a half. It was noticed that only a few students finished the written test in one hour, while the e rest stayed to the last minutes with a handful scores who were reluctant to leave before the right time ran out the examination was held at one time and in one place.

4. Item Analysis of the Entry Test: Discussion and Recommendation

Two important indicators of item performance are to be used, the Facility Index (F.I.) and the Discrimination Index (D.I.).

4.1 Facility Index (F.I.)

The maximum value F.I. can have is (a very easy item and all candidates get it right). The minimum value is 0 (no candidates get it right). It is calculated by dividing the proportion of candidates who obtained the right answer on the number of the candidates who are being tested. In general, the facility indices of the Entry Test were ranged between (0.3 – 0.94). Regarding the reading comprehension questions, question 3 of part 1, questions 1 and 2 of part 2 have facility indices ranged between 0.10-0.19.

Items 1, 5,6,10 of part one and questions 4 and 4 have facility indices ranged between 0.20 -0.29. Question 2 of part one and 3, of part two have facility indices ranged between 0.30-0.39. Question 7 of part one has facility 0.49. Question 8 has 0.51; questions 4 and 9 have facility ranged between 0.60-0.69.

As far as grammar section the facility indices were registered as following:

-questions 1 of part one, 3 of part two; 1, 2, 3 and 4 of part four; 3 of part five have F.I. ranged between 0.0-0.9.

- questions 2,4,6,7,8,9,10 and 11 of part one; 1 of part two; 4 of part three; 5 of part four; 1 of part five; 3 and 5 of part eight have F.I. ranged between 0.10-0.19.
- questions 3 of part one, 2 and 6 of part two, 2 and 5 of part five, 2 of part eight have the F of 0.20-0.29.
- questions 5 of part one, 4, and 5 of part two, 4 of part five, and 4 of part eight have facility of 0.30-0.39.
- questions 7 of part two, 1 of part three have facility of 0.40-0.49.
- questions 2,3 and 5 of part three; 2 and 3 of part seven, 1 of part eight have facility 0.50-0.59.
- questions 1, 2, 3 and 4 of part nine have facility of 0.60-0.69.
- questions 1 of part six, 1,4, and 5 of part seven, 5 part nine have facility Of 0.70-0.79.
- question 4 of part six has facility of 0.80.
- questions 2 and 3 of part is have facility 0.90-0.99 (see tables 2 and 3).

Table 2. Student's marks in admission test

Marks	Frequency	Marks	Frequency
80	2	33	2
70	1	32	1
67	2	31	2
66	2	30	3
26	1	29	3
60	1	28	1
59	3	27	3
57	2	26	2
56	1	25	3
53	1	24	2
51	1	23	4
50	1	22	2
49	1	21	1
48	1	19	1
47	3	18	1
46	2	17	3
45	1	14	1
43	3	13	3
42	2	12	1
41	1	10	1
40	1	3445	91
39	2		
38	2		
37	2		
36	2		
35	4		
34	3		

Mean = 37.8
Standard Deviation = 12.88

Table 3. Facility indices of each question

F. I. %	Reading Comprehension		Grammar Secret Questions							Composition		
	1	2	1	2	3	4	5	6	7	8	9	10
Negative												
0.0-0.09			1	3		1, 2, 3, 4	3					
0.10-0.19	3	1, 2	2, 4, 6, 7, 8, 9, 10, 11	1	4	5	1			5		
0.20-0.29	1, 5, 6, 10	4, 5	3	2, 6			2, 5			2		
0.30-0.39	2	3	5	4, 5			4			4		
0.40-0.49	7			7	1							
0.50-0.59	8				2, 3, 5				2, 3	1		
0.60-0.69	4, 9										1, 2, 3, 4	
0.70-0.79								1	1, 4, 5		5	
0.80-0.89								4				
0.90-0.99									2,			
1.00												

As it is shown in the above table (1) the mean score is fairly small and the standard deviation was 12.88, which indicated abnormal kind of distribution of marks. These statistical procedures and the mean could be interpreted as an indication of the low and poor performance of secondary school learners in English language since the questions of the Entry Test supposed to cover that material.

Stanley and Hopkins (2003) refereed to items with F.I more than 0.40 as very good ones, those within the range of 0.30-0.39 as good ones, those within 0.20-0.29 as marginal and should be reconsidered. But those below 0.19 were very weak and should be rejected.

Henceforth there were, as shown in table 3.

Twenty four weak_questions which should be rejected and not used in any future Entry Test for the low F.I. they have got and it is a good indicator of their difficulty.

These were (question 3 of part one, 1, 2 of part two in the reading comprehension section); questions 1, 2, 4, 5, 6, 7, 8, 9 and 10 of part one, 3 and 1 of part two, question 4 of part three, questions 1, 2, 3, 4, and 5 of part four, questions 3 and 1 of part five, questions 3, 5 of part eight, in the grammar section, it appeared that there were twelve marginal questions, six questions in the reading comprehension, (questions 1, 5, 6, and 10 in part one, and questions 4 and 5 in part two). The other six questions were in the grammar section. (Question 3 in part one, 2 and 6 in part two, 2 and 5 in part five, 2 in part eight). Otherwise there were seven fair questions. These were question 2 in part one, and 3 in part three of the reading comprehension question 5 in part one, question 4 and 5 in part two, question 4 in part eight, of the grammar section.

There were also twenty four very good_questions recommended to be used in any future examinations. Their F.I was above 0.40. From above it is noticed that 8 percent of the reading comprehension.

Questions can be used again in the next again in the next entry examination to the English Department. While only 46 percent of the grammar questions can be used again. Some parts appeared very uninteresting for students such as one and four. Students have seemed very confused with tenses in part one and with the passive voice in part four. This could be explained as tenses were implied in the passive. The reason for their confusion with tenses was worded tin that part. It might be more useful if they were separated.

Three questions (2, 3, and 4) about the use of adjectives appeared to be very easy in the grammar section, part six. Their F.I was 0.90-0.94.

4.2 Discrimination Index (D.I.)

It is a measure of how well a test item can distinguish between the more able (or more knowledgeable) candidates and the less able.

- 1) Arrange all candidates in rank order to their scores (highest first).
- 2) Identify the top quarter and the bottom quarter of the rank (n).

3) For each item determine how many candidates in the top quarter got the answer right (A) and how many in the bottom quarter got the answer right (B).

4) The discriminating index D is then given by:

$$D = (A - B)/n$$

A negative discrimination immediately signifies a faulty question, and the best course of action is to scrap the question altogether. Questions with a discriminating power less than 0.2 are rejected usually by examining boards and those that are between 0.2 and 0.3 are revised. The D.I. of all questions of the Entry Test appears to range between -0.04 – 0.73 (see table 4).

Table 4. Discrimination indices of each question

D.I. %	Reading Comprehension				Grammar Secret Questions						Composition	
	1	2	1	2	3	4	5	6	7	8	9	10
Negative			1(0.08)			(0.04)-4						
0.0-0.09			3				2, 3, 4					
0.10-0.19		1, 2	2, 4		5, 7				3			
0.20-0.29	3, 10	4, 5	5	1, 4, 6, 8, 10, 11	1, 3		1	3	2	1		5
0.30-0.39	1, 6, 7	3		2	4	2	5	1	1, 4	5	2, 3, 5	
0.40-0.49	5				6	3		2, 5		3, 4	4	4
0.50-0.59	4, 8, 9			7, 9					2, 3			1, 3
0.60-0.69	2			3	2	1				2		2
0.70-0.79				5				4	1, 4, 5		1	
0.80-0.89												
0.90-0.99												
1.00												

Two questions appeared with negative discriminating power (question 1 of part two in the reading section, question 4 of part three in the grammar section). These sentences should not be used in any future Entry Test. In addition to that nine questions appeared to have low D.I. these are rejected for their difficulty. These questions are (question 2, 3 and 4 of part two in the reading section; questions 5 and 7 of part 2, questions 2, 3 and 4 of part four, question 3 of part six in the grammar section).

Sixteen questions should be revised as their D.I. was ranged between 0.20-0.29. Three questions were in the reading section (questions 3 and 10 in part one, question 5 in part two); otherwise the grammar section appeared loaded with thirteen questions. Six of them lied within the present and future tense questions (questions 1, 4, 6, 8, 10, 11 in part one; questions 1 and 3 with the definite articles, in part 2, question 3 in part five, 2 in part six, 1 in part seven, 5 in part nine with the reported speech, the adjectives, mass units and prepositions).

Fourteen questions were moderately discriminating as their D.I. ranged between 0.30-0.39. these were: question 1, 6, and 7 in part one of the reading section 2 in part one, question 4 in part two, question 2 in part three, question 5 in part four, question 1 in part five, questions 1 and 4 in part six, 5 in part seven, 2, 3 and 5 in part eight of the grammar section.

Otherwise, sixteen questions had good discriminating power and they are recommended to be used in any future Entry Test in the department of the English.

From above it is noticed that the five multiple-choice questions in the reading section was not promising with one negative discriminating power and three low ones.

Besides three questions out of five in the passive, part 4 had less than 0.08 discriminating power. All these questions are rejected.

Finally, view the average percentages of the discriminating power of all part of the Entry Test, it is noticed that the passive voice sentences (0.16) and the multiple choice questions (0.11) had low discriminating power. Therefore, these two parts should be reconsidered carefully, possibly replaced by better sentences. As a whole

the reading comprehension section (0.28) appear with a lower discriminating power than the grammar one (0.36).

5. Conclusion: Few Remarks about the Multiple-Choice Questions

The low discriminating powers with one negative of the multiple choice questions in the second part of the reading section might be rooted back to the confusion the test was in. It is noticed that:

- 1) While the right answer was D. To question 1, a large proportion of students answered B. And just the least answered correctly. This could be reasoned to that B might be considered more logical to students (more people attended because it was held earlier).
- 2) Student did not choose the right sentence to question 2 (it was C) the swung to D. Here it might be related to what they would like to do if they were asked to do the same thing (if the exam is difficult, it is better not to take it).
- 3) In answering question three, some students considered D closer than B which is the right answer. Here it might the simplicity of this sentence was a good indicator. Why they chose D rather than the right one? A large proportion left this question unanswered.
- 4) It was unusual to find that the same number of student chose A and C while for the right answer, it was a little bit less than that.
- 5) Sentences A and C appeared very confusing to most students. They could not understand the meaning of “unlikely” properly some of them decided to leave this question unanswered. (See table 5). This question, with the low discriminating power and low facility index, must be rejected. Students appeared untrained in this type of question especially if we take the traditional method of learning English in school into consideration. The multiple-choice questions were very good techniques for assessing students’ reading comprehension but it might be more advanced to them. The difference between the discriminating powers of the traditional reading passage flowed by the ten Wh questions and yes/no and the multiple choice questions showed that they knew more about the first teaching method than about the second one.

Table 5. Students' answers to the multiple-choice question/part two in the reading section

Questions	Right Choice	Choices					unanswered	Total
		A	B	C	D			
1	D	14	36	19	18	4	91	
2	C	14	24	16	33	4	91	
3	B	6	28	12	37	15	91	
4	D	23	17	23	21	7	91	
5	A	22	27	24	11	7	91	

Regarding the composition, in examining students’ responses, it appeared that 30 students did not answer the letter writing question and 18 students left the paragraph writing unanswered. It was noticed that most students faced problems in writing proper English, ranged from mistakes in grammar to spelling and punctuation, the obvious weakness in writing skill might due to the lack of knowledge and training in writing as a reasons of negligence of composition secondary school level.

References

- Brown, H. Douglas. (2003). *Language Assessment - Principles and Classroom Practice*. London: Longman.
- Harris, D. (1969). *Testing English as a Second Language*. New York: Mcgraw- Hill.
- Manish. A. Vyas & Yogesh L. Patel. (2009). *Teaching English as a Second Language: A New Pedagogy for a New Century*. Prentice Hall India.

- Oller, J. W. (1979). *Language Tests at School*. London: Longman.
- Oller, J. W., Jr., & Perkins, Kyle. (Eds.) (1978). *Language in Education: Testing the Tests*. Rowley, MA: Newbury House.
- Rivers, W. M. (1968). *Teaching Foreign Language Skills*. Chicago: The University of Chicago Press.
- Rountree, D. (2008). *Assessing Students: How Shall We Know Them*. New York: Routledge.
- Rowley, G. (2005). The Relationship of Reliability in Classroom Research to The Amount of Observation: An Extension of The Spearman-Brown Formula. *Journal of Educational Measurement*, 15(3).
- Scriven, M. (1967). The Evaluation of Methodology of Evaluation. In R.W.Taylor (Ed.), *Perspective of Curriculum Evaluation*. Chicago: Rand McNally.
- Scriven, M. (2010). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Chicago: Rand McNally.
- Stanely, J. C., & Hopkins, K. D. (2003). *Educational and Psychological Measurement; and Evaluation*. New Jersey: Prentice-Hall Inc.
- Sutton, R. A. (1985). *An Introduction to Assessment and Evaluation: Processes and Procedures*. Cardiff: Faculty of Education, University College.
- Valette, R. M. (2000). *Modern Language Testing*. New York: Harcourt Brace Jovanovich.
- Youngman, M. B. (1979). *Analysing Social and Education Research data*. Maidenhead: Mcgraw – Hill.