

Evaluation of Language Testing: An Applied Linguistic Perspective

Sadeq Ali Saad Al-Yaari (Corresponding author)

Independent Researcher, Dept. of English, College of Arts, King Saud University (KSU)

Riyadh, Kingdom of Saudi Arabia. E-mail: prof.sadeq@gmail.com

Fayza Saleh Al Hammadi

Associate prof., Dept. of English, College of Arts, King Faisal University, Al-Hassa

Kingdom of Saudi Arabia. E-mail: viceii.library@kfu.edu.sa

Salah Ayied Alyami

Assistant professor, Dept. of English, Dammam College of Technology, Dammam University

Riyadh, Kingdom of Saudi Arabia. E-mail: salah1sa@yahoo.com

Received: January 1, 2013 Accepted: January 10, 2013 Published: January 17, 2013

doi:10.5296/ijele.v1i2.3059 URL: <http://dx.doi.org/10.5296/ijele.v1i2.3059>

Abstract

Background: Over the past few years, researchers started to pay great attention to studies on language testing to improve the methodologies of administering, constructing and evaluating tests.

Aims: The present study aims at providing a new perspective as it contributes a new method based on scientific analysis that does not only diagnose the phenomenon, but also provides solutions and suggestions.

Methods: The study was conducted at the Departments of English both in the College of Education and College of Arts & Linguistics, Tamar University, Yemen. Ten (10) final term tests administered by ten instructors within two colleges were evaluated. The tests were given to freshmen students to test their productive, receptive skills, as well as their language components.

Conclusions: Results show that some tests lack validity, practicality, reliability,

discrimination, accountability, and wash back effect. Findings also show that instructors' rubrics were neither clear nor brief. Questions also involved orthographic as well as grammatical mistakes. Finally, in comparison with instructors of College of Education, instructors of College of Arts are found to be better in administering tests.

Keywords: Language Testing, Evaluation, Assessment, Errors, Applied Linguistics

1. Introduction

Large body of research has been done on issues related to language testing and their developments throughout the past few decades (Daviers, 1968; Simpson, 1972; Clark, 1972; Johansson, 1973; Johansson, 1975; Kulakow, 1976; Ipswich, 1978; Adams, 1978; 1979; King, 1981; Kohonen, 1988; Harley, 1990; and Heining-Boynton, 1990). Nowadays, it is almost agreed upon most linguists in general and those involved in language testing in particular that test should be considered from two perspectives: One theoretical and the other practical. The theoretical perspective involves issues related to the requirements of designing language tests, typology, characteristics, etc. (Bachman, 1990) and practical aim including test evaluation, designing test in the following areas: Testing sound system, testing language structure, testing vocabularies, and finally testing language skills.

As tests have traditionally been the exclusive domain of the teachers, many treat the issue of what do we want examinees to perform including language for survival needs e.g., eating, direction, hotel, etc. For eating for example, we need to know something about the literature of food, analyze it, etc. Some discussed the idea related to what extent testers can benefit from the results of previous language tests (Anivan, 1991). For instance, McNamara (1991) investigated the role of Item Response Theory in Language Test Validation in Australia. One of the outcomes of the study is that the test is indeed one-dimensional, and supports the use of Item Response theory for such analysis. Another important result is that the kinds of listening tasks in the two subtest parts represent significantly different tasks in terms of ability required to deal successfully with them.

Comparing the analytic with holistic methods in matters concerning scoring methods, results of Gilfert & Harada (1992) show that both ratings were very close, with the maximum difference of three points in a 20-point scale, and the ratings had a high correlation with the writers' TOEFL scores. Such results, however, were later supported by Washington (1992). It is felt that a distinction ought to be made between assessment and evaluation though some researchers are object to such differentiation and this what has been proved when investigating the tests administered in the American colleges and universities (Al-Ghamdi, 1993; Brwon (1996). This motivated some researchers to evaluate cross language system as it, according to the holders of this viewpoint, provides an infrastructure for the testing and evaluation of information retrieval systems operating on European languages (Peters, 2002).

Linguistically speaking, English language skills have been differently assessed by many researchers. Examining the effects of reading on understanding proficiency tests including TOEFL, Jang's research (2005) demonstrates that the purpose of any skill in general can be more effective when it achieves its purpose. Lassche (2006) discussed the usefulness of feedback framework of the test including reading. He investigated the phenomenon in terms of Developing and Evaluating Procedure for Describing Measures of Student Achievement Employed in Web-Based Language Learning Environments. According to the Lassche, the program evaluated shows little validity with its stated objectives, tending towards informal, pre-communicative assessment that reduces the status of teachers and frames student performance in terms of discrete language functions.

Generally speaking, assessment and evaluation are roughly synonymous (Flucher & Davidso, 2007). So, the question now is the following: How testing and assessment are related to evaluation? To answer this question, we need to go back to the early definitions of the topic where one can find that testing and /or assessment are considered a part of evaluation just like questionnaire, observation, etc. (Harrison, 1980) It is for this reason, however, that this study has been called an evaluative study and not an assessment one. Tests selected for this study are artificial 4-skill tasks wherein as we will see, each test is made for one skill. Reading skill is one of the most difficult tasks that normally face students who want to undertake proficiency tests like TOEFL, IELTS, entrance test, admission test, George Town test, Michigan test, etc. It is for this reason, however, that it has been broadly discussed in many languages (Alonzo, 2011).

Mercier (2007) has investigated the effects of one of the professional programs called "Focused Approach to Frontloading English Instruction"(FA) on the writing of advanced learners of English. The researcher concluded to saying that the program has an effective impact on the structure of L2 writing levels. Comparing the effects of the overt letter-sound training to holistic word training on the development of word recognition in the orthographic paradigm of a novel, Bolger (2011) concluded that this strategy is useful as it leads to optimal word recognition performance in the system of alphabetic writing because of its mechanism. Based on previous studies done in light of the scores reported from the National Assessment of Educational Progress (NAEP), Dennis (2011) has concluded from his multivariate correlational study on young strugglers' adolescents by indicating that one-size-fits-all approaches to policy and instruction relating to struggling young adolescent readers do not meet the heterogeneous needs of this population of students.

Raynold (2011) investigates the differential acquisition of English letter-sound correspondences in bilingual children. Outlined results of the study show that there is no significant difference between the monolingual group and the bilingual group on the non-word test. Language tests were used for specific purposes including medical fields like using it for dyslexia assessment (Lay Wah Lee, 2008). Moreover, some researchers decided to estimate language tests in their countries. For example, Qian (2008) investigated Hong Kong most important tests as it might help them in making new testing and assessment procedures in the regions. Ramanathan (2008) provides a description of the forces and factors that define educational structures in India and the political and academic responses to language issues. Inn-Chull (2008) discussed the situation in Korea. The focus here was on the impact of EFL testing on EFL education. One of the findings shows that a number of young students are under pressure to take the EFL tests. Liying (2008) investigated the phenomenon in China. The researcher looks at English language testing as the key to success. Sasaki (2008) discussed the same issue in Japan in the light of the description of the 150-year history of the school-based English language assessment.

Analyzing the reasons that might be behind the believes and practices of the elementary school teachers when they teach writing skill, Glen (2011) concluded from his studies that those involved in pedagogy should learn more about who is the teacher and what is he/she supposed to do in the class, especially in matters concerning writing. According to the

researcher, such procedure will help pedagogues move toward authentic literacy and science writing practices. Exploring whether there are relationships associate reading skill as an interest with some other interests and/ or skills whether be it general or be it personal, the results of West (2011)'s study show that reading interests are exclusive of general life interests and personality dimensions. However, because of its importance of correlation studies, they have been used by many researchers. Attempting to understand the relationship between below-grade-level reading and graduation rates in one of the school systems in George by comparing their advantages to their disadvantages, Robinson (2011) concluded that there is significant relationship between reading ability and students' ability to finish the requirements that will qualify them to receive their high school diploma successfully.

Massar (2011) examined a case study using the Corrective Reading Program in a junior/senior high remedial class, and found that the achievements and level in student's performance in reading skill had improved after they enrolled in the Corrective Reading Program. The results of this program can be clearly seen through the difference between the results of pre-test and post-tests. As studies improve with the passage of time, researchers try their best to find some strategies and /or technique that might be useful in teaching language's skills or even one of them. We can see that in Fergus' study done in 2011 that investigated the effects of using graphic organizers on writing summaries in the class and concluded to saying that students' ability in constructing such kind of activities and interventions has significantly improved. The same strategy has been used in 20011 by Esmat, but this time along writing checklist, resulting that students' scores have increased positively. Discussing how writing can be used to foster students' metacognitive skill improvement, Davis (2011) concluded that eight participants assigned writing were both expository and imaginative and ranged in frequency from once a week to once a year. In fact, most of these techniques were finally directed to qualifying students to proficiency tests (Heckman, 2011). This direction determines understanding the theoretical developments in validity research in L2 performance testing (Uysal, 2009).

Assessing third grade English students in reading and English oral language development, Acosta (2011) concluded that English oral language accounted for approximately 22-30% of reading performance variability. Additionally, bilingual education instructional models (experimental and control) experienced the most growth in English verbal and listening comprehension skills compared to the growth rate of the SEI-T group. Some researchers tend to draw a cause/ effect relationships between reading skill and other factors like the impact of orthographic conventions and morphological features on reading (Roberts Frank, 2011) the teacher (Lawson, 2011), the phonemic factor (Walter, 2011). This type of researches covered all members of society including children in kindergarten (Chapman, 2011).

Some decided to measure, assess, and evaluate language testing in higher education (Secolsky, 2011). According to these researchers, measurement, assessment and evaluation requires fully understanding of the developments of language tests even by meeting experts in the field and those decision makers. Chen (2011) for example interviewed Professor Lyle F. Bachman, the former president of the American Association for Applied Linguistics and of the International Language Testing Association. One of the outcomes of the interview is the concentration of

the factors that affect language testing like the distinctive social and cultural values that characterize the traditions of language assessment in the East and the West. Some suggested making use of new methods for alternative assessment and evaluation (Rahman, 2011). Comparing Traditional versus Recent Trends in Evaluation relating to Language Teaching, Gautam (2011) concludes that modern evaluation system is better in comparison to older one for it attempts to measure a comprehensive range of objectives of curriculum. Investigating language testing's evaluation as an important aspect of the process of language learning and teaching, Rahman & Gautam (2012) show that linguistics can be of great help in developing the methodology of constructing language tests.

1.1 Aims

Identifying the mistakes made by Thamar university's two instructors in terms of how and why they have committed such mistakes is one of the aims (objectives) behind undertaking such study. The researchers also attempt to present model of how the test should be conducted. More specifically, the researchers attempt to answer the following questions:

1- What are the most common mistakes committed by university instructors? How can we recognize them? What are the best ways for evaluating them?

2- The results of the analysis. To what extent can they be useful for a student in proficiency tests and other general tests?

1.2 Methodology

The researcher has used ten (10) tests made by 10 university instructors teaching at the College of Arts& Linguistics (officially known as Arts) and college of education (officially known as Education) in the academic year 2009/ 2010. These instructors are: Arts' Grammar Instructor (AGI), Arts' Listening Instructor (ALI), Arts' Reading Instructor (ARI), Arts' Speaking Instructor (ASI), Arts' Vocabulary Instructor (AVI), Education's Grammar Instructor (EGI), Education's Listening Instructor (ELI), Education's Reading Instructor (ERI), Education's Speaking Instructor (ESI), and Education's Vocabulary Instructor (EVI).

The exams were made to test three language's skills: productive skill (Speaking), and receptive skill (Reading and Listening) and two language component exams: Grammar and vocabulary. Note here that writing skill is not involved, because it is not involved in the two institutions' curricula. Instead, they have a syllabus called vocabulary instead.

Having data collected, the second step was to identify these data. The processes of identification were followed by fully classification and complete description and explanation. The researcher then started the process of rating these data by evaluating them in terms of how/ how not and why / why not instructors have made such mistakes before these mistakes improve to become errors.

2. Results & Discussion

2.1 Tests of College of Arts & Linguistics

2.1.1 Grammar Test

In his first question: "Read the following sentences and write (T) for the correct ones, and (F) for the wrong ones with correcting the wrong ones", Arts Grammar Instructor (AGI) makes a mistake when he used (T) and (F) to mean correct and incorrect (wrong) as the verb "write" here shows that it is followed by a word, not by a letter (F) or (T) that are grammatically considered as "Ticks" than "words". Another mistake is that when he adds the preposition "with" which cannot be used here as we can just add the comma followed by the noun gerund "correcting". Hence, the question should have been written as follows:

"Read the following sentences and write (correct) for the correct one, and (Incorrect) for the wrong one, correcting the wrong ones:"

AGI makes another mistake when he writes the word "some" in the fourth item of his question when he writes: "He promised her to some on time." It seems that what he wants to write is the word "come", not "some" which might be seen as a lexical mistake, but the fact is that the word "some" has something to do with grammar (especially when it is followed by the word "time") and that what makes it a grammatical mistake rather than lexical one.

One more example can be evidently seen in the fifth question when AGI writes: "Read the following sentences and choose the correct word (linker) for each of the following sentences." as he could simply replace the word "Linker" with "linking word" so that the correct sentence would have been written as follows: "Read the following sentences and choose the correct word (linking word) for each of the following sentences." Additionally, the way he puts the item is not professional, because he puts them in boxes, while the fact is that Multiple Choice Question (MCQ) should be used regularly and in classical way.

Generally speaking, the whole test is good. It matches the level of the students. It tests students' recognition of the present, past, and future tenses in their simple forms with the use of indicators (clues). It also tests their recognition of the sub-verb agreement. Finally, it tests the ability of the students to form correct questions from statements (transformation, especially with non-helping verbs which are problematic to some learners.

2.1.2 Listening

In his fourth and fifth question, Arts Listening Instructor (ALI) does not set it clear for us. Consider:

4. Listen and write what you hear (tape)

- a. Part.
- b. Ear.
- c. Water.
- d. Stair.
- e. Hear.

5. Listen and write down what you have listened (tape)

- Ten men.
- This shape.
- This unit.
- Would you try?
- Red pen"

It is not clear how students are going to be tested his student in listening skill by ALI. Is he going to give them these two questions in particular as a "dictation"? There is no comprehension involved. Moreover, question five (5) seems to be a question on sound system, not listening comprehension. It is not clear how ALI is going to administer his test. Put differently, we, as readers, do not know how the test is done. Do all students follow the same instructions? We know that distinguishing testees is one of the characteristics of the good test called (discrimination).

There are some other mistakes in the test of ALI, but not in listening comprehension. Some of them are classified as grammatical mistakes as in his second question when he writes: "Q2. Listen to the following short dialogue identifies a participant in each dialogue according to the instructions given with each dialogue: (tape)." which lacks a co-coordinator as can be clearly seen. To make it correct and meaningful, ALI should have written "and then identify"; otherwise, the question will be understood as "The dialogue identifies...."

Another similar example, but this time with relative pronouns where it is found that he omits "who" in the second sentence in the fifth section of question number two when he writes: " 2. That he feels sorry for the speaker(A) does not have a borrowing card (membership card)." whereas the correct sentence should have been written: " 2. That he feels sorry for the speaker (A) who does not have a borrowing card (membership card)."

Another mistake is that when the rule of present simple tense is not correctly applied as can be seen in the third section of question number two when he writes: "Q. What does the speaker (c) means in last utterance?" wherein the sentence should have been written: "Q. What does the speaker (c) mean in last utterance?" without the inflexional suffix "s". Despite it is a grammatical mistake, it is considered a demerit in the exam, especially when it comes in the question which, in turns, may affect student's answer to the main topic. .

Mistakes on Prepositions were also found in the test of ALI as can be seen in his sentence when the preposition "in" is added as can be seen in his sentence: "He is a teacher in the college where in these students are going to study." As here the addition of the preposition "in" is unnecessary, the sentence could have been written: "He is a teacher in the college where these students are going to study." Spelling mistakes are also existing as can be seen in the fourth section of question two when he writes: "I am really so much sorry." whereas he should have written: "I am really so much sorry."

Last, but not least, it is found that ALI commits mistakes in punctuation marks like when all statements in the third section of question two are put without full stops e.g., "He is going to

be their friend" or when he adds question marks to affirmative sentences as in his third sentence in section one of question number three when he writes: "I want to have my hair cut?" and also as in his third sentence in the fourth section of question number three when he writes: "I want to borrow this book please?".

ALI fails also to use commas as in his first sentence in third section of question number three when he writes: "What is the matter children? as we find him omits the comma which should be written before the word: "children" so that the sentence would be: "What's the matter, children?" Eventually, ALI misuses adverbs like when he writes: "I am really so much sorry." wherein the adverb "much" should be omitted as it cannot be used here.

Overall, the test is not clear. The instructor could test his students' ability to understand given instructions for performing class task. The test also tests students' ability to comprehend basic dialogue in English.

2.1.3 Reading

Arts Reading Instructor (ARI) makes a mistake when he puts a very long question. The testee will forget the beginning of this phrase when he comes to the end of it. This can be clearly explained in view of his question when he writes: "Read the following passage and then decide: What is the most suitable choice from among the given choices; can be the main idea for the passage you have read?"

There is more than one question in this question. By the end of the question, the testee will forget the trigger. Also, the use of the semi colon makes it more difficult to be understood. Despite the fact that this might be considered an orthographic mistake and one might claim that it has nothing to do with our analysis which is supposed to focus on reading, we cannot separate that from our analysis as such mistakes may mislead the testee.

Question one has too many options. ARI writes six (6) options which make it unreliable and also impractical. Consider:

"1-What is the best title for this passage:

Football teams.

Playing football.

Playing football in India.

Playing football in Kingdom of Saudi Arabia.

Saudi and Indian teams and their supporters.

Watching football matches."

In question three, ARI asks the questions without giving students answers to select from. It seems that he also has a problem with punctuation marks, because he misuses them almost in all his questions. In fact, it seems that ARI cannot distinguish interrogative sentence from imperative one. It is for this reason that we find him, for example, using question marks at the

end of all imperative sentences throughout the test. Example: "Answer the following questions, choose the correct answer for each one?" which makes it necessary for him to rephrase them so that the former question, for example, is written: "Answer the following questions. Choose the correct answer for each one."

ARI also found to be unaware of the way how the questions are made. We find him repeat the question more than one time as in his third question when he writes: "Read the following passage and answer the given questions below?" and again repeats the same question in a different way right after the passage is finished as follows: "Answer the following questions with 'Yes' or 'No'?"

In his fourth question, ARI commits a serious mistake as he asks a question in vocabulary for reading comprehension. This is better being explained in light of his question when he writes: "Find words from the above given text which matches the following pictures. Write the word under the picture directly, the first one has been done for you.?" Of course there are some insects' pictures which show it is suitable for reading comprehension not vocabulary.

Other mistakes include the misuses of spelling as in the word: "Holy" in his second question which lacks another "L" and the word: "television" which has to be in plural to agree with words before and after it. One more mistake is that when he replaces the word "others" with "other", and the preposition "at" with "in" talking about the job and it is known that the preposition that goes with the occupation is "at". Consider his sentence: "A student will study in any one of the partner European universities." Another mistake in the same sentence is that when he adds the word "one" which converts the usage of the word "one" from a pronoun "any of the partner European universities." into a modifier "any one of the partner European universities." Again, we cannot weight much upon such mistakes as they are not directly related to the aim goal of the analysis which is concentrated on the mistakes on reading comprehension.

Finally, it should be noted here that the good aspect about the test is that it could test students' ability to read a text and comprehend the main idea (scanning). It could also test their ability to comprehend a text in details.

2.1.4 Speaking

The first question of Arts Speaking Instructor (ASI) is not clear. Consider:

"Q1. Underline the target sound(s):

Beg	Pig-Big-Bag.
Vat	Fat-Fate-Van.
Three	They-Thanks- this.
Chair	School-Teacher-Kitchen.
She-	sheep-See-Ship."

As can be clearly observed, testees cannot understand what their tester wants them to do.

Does he want them to underline sounds that are similar to the trigger ones in the first column? If so, then why does not he underline them? Besides, the first choice in the last sentence is not suitable. The word: "sheep" should have been substituted by a word like "shift" for example due the similarity between this sound and the target sound.

The second question is vocabulary implied. Consider:

"Q2. Select a picture from column B and match it with words from column A.

Pin A picture of a pen/ a picture of a pin."

In the third question, ASI commits a serious mistake as the whole question fits testees of vocabulary comprehension, not testees of speaking skill. Consider:

"Q3. Fill in the blanks to complete the story

Ahmad

Once there was a...  named Ahmad. He was

years old. Last week, he and Ali rode on a... 

to school. He saw a...  on the street. He walked in

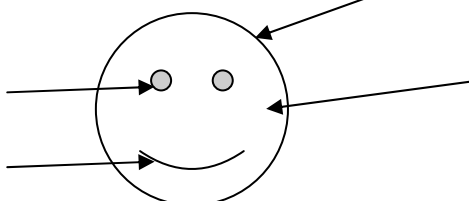
the  of the school and rode up to the 4th floor on

the  . He looked out the...  and saw

the...  on the street. He has good  ."

The same problem repeats itself in his fourth question. The question does not test students' ability to recognize sounds or to produce them or at least discriminate them. Conversely, it prompts testees' ability to write. Consider:

"Q4. Look at the facial diagram and complete the words:"



However, the purpose of the items of question five varies from grammar to vocabulary. Consider:

"Q5. CHOOSE BETWEEN BRACKETS:

Whose book is this? It is(mine-I-myself).

Fadi is Mr. Saleh's son. Mr. Saleh is Fadi's (brother- father-uncle).

The nose is (inside- on- under) the face.

The last month of the year is (April-October-December).

The word that starts and ends with "P" in the following is:

(Pen- Lamp-Peep)."

Obviously, the first and third items are considered grammatical rather than speaking skill questions, while the second and fourth questions are vocabulary questions (lexical items) which make the four items suitable for language components, not speaking skill. Additionally, the fourth question, ASI commits another grammatical mistake when he adds the central determiner "the" (Article) before the word "last" (adverb) and we know that it is impossible to pre modify such adverb. The fifth item is not clear. ASI should have remembered that he is testing speaking skill, which means that after recognizing the sound, student is supposed to be asked to produce it (Oller & Perkins, 1978).

Eventually, it should emphasize on the fact that the instructor could achieve some of his objectives. For example, he could test students' ability to discriminate and recognize minimal pairs e.g., (P.b), (F.v), etc.

2.1.5 Vocabulary

In his first question, Arts Vocabulary Instructor (AVI) makes a mistake by using the word "Bon mot" which is not English word (French word). However, this is wrong as we test students' performance in English, not French as such it violates one of the characteristics of the test (validity) where the test measures what is supposed to measure. The fact that the question is True/ False (T/F) means that the guessing element is high (50% for T/F, 33% for T/F/DK). Consider:

"1- 'Bon mot' is the synonym of 'clever mark'."

The second question, another mistake is also found when AVI uses wrong options. Consider:

"4-A vendor must have a license.

A car driver (b) anyone who sells food. (c) anyone who works in a hospital
 (d) emergency patrol."

However, it is known that wrong distracters in MCQ can distract good students. Moreover, the second option is totally wrong, because the closest meaning to the word "vendor" is anyone who sells things be they food or any other object. Additionally, such questions need high guessing which we know this characteristic in them is one of their disadvantages.

The third question of AVI does not test vocabulary at all; rather, it is appropriate for reading

comprehension test. Consider:

"Q3. For each of the following, choose the answer that is closest in meaning to the original

The moral of the story is clear" Don't kill the goose that lays golden eggs."

(a) Every story is clear:" Don't kill the geese that lay golden eggs.

(b)Don't trust appearance, as they deceive us easily.

(c)The conclusion of the story is easily seen" Greed does not pay"

(d) Don't believe the proverb "all that glitters is not gold."

In his fourth question, AVI uses words that are not familiar to students. Such procedure is incorrect as we are supposed to measure the performance of the student based on what we have given him, not what we have not or what we expect them to do. Again this makes the test invalid. The picture can be clarified more in view of his question. Consider: "Q4. Underline the odd word of the following:

Aardvark- Badger-Armadillo-Cockatoos.

Ostrich-Cracker-Cricket-Sea gal-Dove.

Limes-apricots-Parsnips-Lychees.

Buff-magenta-lilac-oval.

Whirr-Jingle-Zip-Fizz."

Furthermore, the tests are supposed to construct our students' ability, not destruct ambitious spirit in them (Henning, 1987; Heaten, 1988). In other words, we should tests what students have studied, not try to catch him out in what he did not study (Jacobs, 1981; Madsen, 1983; Hughes, 2002) as this will affect the wash back effect and cause it negative impact on both learning and teaching (Oller, 1983).

As a matter of fact, the picture is not that dark. There are some advantages in this test. The instructor, for example, could test students' abilities to produce plural, past and (ing) forms for some confusing words, it tests production. Some questions have double folded; they test the students' ability to recognize words according to their classes (parts of speech) and also words in decided to their meanings (antonyms). Finally, the test could examine the students' ability to recognize the different usage(s) of some words which sounds similar, but used differently.

2.2 Tests of College of Education

2.2.1 Grammar

It is found that Education Grammar Instructor (EGI) does not know how to write questions. His third question is incomplete. Consider:

"Q3. Transfer the following sentences."

problematic to some learners. One more merit of it is that it tests students' ability to use linking words in connecting sentences.

2.2.2 Listening

Education Listening Instruction (ELI) seems to have a problem in making questions. In fact, the test seems to be made for testing the speaking skill not listening skill which makes it invalid. However, the questions are not even correctly formed. Looking at his first question, for example, we find that ELI could not construct it well. Consider:

"1. Introduce yourself taking into consideration answering the following questions:" wherein the right way of writing that question is as follows:

"1. Introduce yourself by answering the following:"

The same problem occurs in the third question when he writes:

"3. Why you have joined the department of English? You should listen to this question. Answer it." while he should have written it as follows:

"3-Why have you joined the English Department? Listen to this question and answer it."

Another problem in the question is that of the misuse of punctuation marks mainly capital letters wherein ELI fails to use them as can be seen in the word:

"Department" which he wrote its initial letter "d" in a small letter.

In the fourth question, ELI misuses the verb "write" by replacing it with the verb "take" thinking that the noun "note" takes the verb "take" as can be seen the question. Consider:

"4. Choose one of the following topics/ issues/ items to talk about it, take notes if you would like:"

However, in his fifth question: "5. Discussion on the basis of the above chosen topic by the student?" we find that ELI does not make it clear. Firstly, using the word "discussion" here in this position is inappropriate. Secondly, since we have many topics mentioned, the word "topic" used by ELI in the question is incorrect. However, he should have pluralized it. Thirdly, ELI ends the statement by a question mark as if it were a question whereas it should not be done because it is a statement, not a question.

Finally, one needs to know "how" he is going to make use of the above mentioned topics he wrote about. It is not known how the questions and/ or items mentioned above are going to be employed. ELI forgets that students do not have tape recorder to listen to the tape talking about any of the above mentioned topics. Therefore, it is unclear whether he is going to raise a topic based on any of the above mentioned items and then ask them to answer questions, or he is going to provide them with pictures, for example, and then again ask them to match them while narrating a story based on one of those items.

All in all, when accounting for the test's questions, it can be said that they test the students' ability to recognize and infer, through listening (if a tape recorder is used) the situations

(locations) of the listened dialogue and / or conversation. The test also tests students' ability to recognize words pronounced with certain sounds, and also to reflect what they have heard productively (with the use of writing of course).

The test focuses on assimilation, problematic issue for the majority of non-native speakers of English. The five given items are pronounced differently (final sound of each first word and initial sound of each second word), from the way they are written as a result of assimilation.

2.2.3 Reading

The same error made by ELI occurs by Education Reading Instructor (ERI) who tests his students in reading comprehension by giving them a writing test. This makes the test. In fact this was not the only mistake. There are some other mistakes related to the questions themselves. For example, we find him adds unnecessary words like when he writes:

"Q1. Answer the following questions by reading them and then writing either a simple or compound sentence for each according to the questions?" wherein he should not have added the prepositional phrase: "...by reading them and then..."

In the third item of the first question, ERI fails to use the right auxiliary verb. He replaces the verb to be "are" with the verb to do "do". Consider:

"C: How many brothers and sisters do you have and do your parents still alive?" whereas the correct sentence should have been written: "C: How many brothers and sisters do you have and are your parents still alive?"

More importantly, a single question should not ask for information about more than one item. The same problem of the former question occurs in this question when he adds the unnecessary preposition: "with" as can be seen in his question:

"Q2. Read the following paragraph and rewrite it with correcting the mistakes and errors?" where he could have simply adds a comma and then start the second part of the sentence by the word: "correcting". Consider:

"Q2. Read the following paragraph and rewrite it, correcting the mistakes and errors?" In fact, ERI could have ended the question by the word: "mistakes" as the sentence will be complete that way. Conversely, adding more words to the rubric makes it unacceptable as clarity and conciseness are two necessary conditions in the instructions. Additionally, there is a difference between "mistakes" (performance, unsystematic, irregular, self- corrected by the learner, etc.) and "errors" (competence, systematic, deviant, inter language competence, are not/ cannot be dedicated/ corrected by the learner, etc.)

ERI has a problem with punctuation marks mainly question marks. He adds them at the end of all his questions without exception while the fact is that he should not have done that as the questions have been written in a form of affirmative sentences, not interrogative ones. Consider:

"Q1. Answer the following questions by reading them and then writing either a simple or compound sentence for each according to the questions?"

Q2. Read the following paragraph and rewrite it with correcting the mistakes and errors?

Q3. The following lines (verses) have been written by someone as his dream, read the carefully and try to write your dream in the same way?

Q4. Read the following announcement and write an application letter; that is, a response letter indicating the following points?

Q5. Think of any topic that you read at once then write about it, write first the title and then write a paragraph?"

As can be seen, ERI misuses the punctuation marks that he uses them randomly in all his questions. This randomization in using punctuation marks contradicts the rules of writing. It also goes against the rules of instructions which require that punctuation marks should clear and brief. The situation becomes worse if we know that this is against the general rule says: "understanding the question means half of the answer".

Ultimately, positive aspects of this test should not be ignored as the instructor finally succeeds to test students' ability to read and predict meanings of difficult words through reading; that is, words in context. It could also test students' ability to read and comprehend what they have read so that they act accordingly.

2.2.4 Speaking

When he designs his first question, Education Speaking Instruction (ESI) makes a mistake by writing the target words which should be concealed from students. The reason for that is so simple: It is a spoken test and testees are supposed to listen to a tape recorder, recognize the sounds and then respond (Underhill, 2005). Consider:

"Q1. Listen to the following pairs of words and decide if their initial sounds are similar or not (Use (S) for similar pairs and (D) for different pairs, the first one has been already answered as a sample?

<i>A</i>	<i>B</i>	<i>Answer (S or D)</i>
Pen	Ben	D
Cap	Cab	
Pin	Bin	
Big	Pig	
Pad	Bad	
Pray	Bray	

In addition to his mistake of making the instructions very long as can be seen in the above mentioned question, ESI commits another mistake by failing to write long numbers for the items. Also, there is another mistake related to rubrics. The instructions should be rewritten as follows: "Q1. Listen to the following pairs of words and decide if their initial sounds are similar or not (Use (S) for similar pairs and (D) for different pairs, the first one has already been done?

The same problem of numbering repeats itself in the second question. However, there is one

more problem in the second question related to the context. It may give a clue to the correct answer.

Consider:

"Q2.Tick (/) the words you recognize in the sentences you hear(the first one has been answered as a sample):

<i>A</i>	<i>B</i>
Fine	Vine
Few	View
Leaf	Leave
Half	Halve
Fan	Van
Feel	Veal

-I got beautiful flowers and vine.

-This writer's view is negative.

-None will leave the class till we finish the exercise.

-I spent an hour and a half for this assignment.

-Veal is a very tasty meat."

There are more mistakes in the items of the previous question themselves. The verb "tick", for example, should be replaced by the verb "check". The pronoun "none" should be replaced by "no one" and the place adverb "till" should be replaced by "until". The preposition "for" should be replaced by the preposition "in". Finally, the word "meat" in the last item is uncountable noun. How can it be modified by the indefinite article "a"?

The mistake committed by ESI in question three is related to the type of the test itself. This can be clearly understood through the third question. Consider:

"Q3.Listen, repeat and write down what you hear:

Ten men-this shape- this unite-would you try-red pen."

We need to ask ESI what he means by the question. The researchers believe that the test is not a speaking test anymore. Rather, it is a spelling test which makes it to some extent invalid. There is another problem related to other characteristics of the test including reliability, practicality, etc. As can be seen here, the test is short which makes it unreliable and unpractical either.

Nor must we forget here to notify that the instructor could test students' ability to produce phrases with assimilation. In fact, the test examines students' ability to both discriminate and recognize many sounds with different spellings and graphic representations.

2.2.5 Vocabulary

Education Vocabulary Instructor (EVI) first mistake is that he substituted the verb "circle" with the verb "identify". More importantly, he wants testees to identify the odd (different) word in each group, but he does not explain how they are supposed to perform this procedure. Consider:

"2. Read the following words and identify the odd word in each group:

Teacher	Lecturer	Tutor	professor	father	trainer
Song	Music	jazz	Singer	Cassette	Preacher
Student	Learner	pupil	Researcher	headmaster	
Moon	sky	sun	son	space	
Head	Touch	Smell	See	Speak	feel

EVI's third question was another area of committing mistakes. In one time, he asks the testees to tick the answer and in the second time, he asks them to underline it. Consider:

"3. Answer the following question according to the given instruction in each item:

One of the following words is not a noun, underline it:

God	Good	Mood	Food		Wood
-----	------	------	------	--	------

b. One of the following words is not a verb, tick it:

Enlarge	Emphasis	Wait	Simplify		Minimize
---------	----------	------	----------	--	----------

c. One of the following words is an adjective, tick it:

Badly	Easily	Manly	Certainly	Simply
-------	--------	-------	-----------	--------

d. One of the following words is an adverb, tick it:

Now	Womanly	Lovely	Hello	This
-----	---------	--------	-------	------

e. One of the following words is the opposite of the word wrong, underline it

Bad	Good	Right	Write	incorrect
-----	------	-------	-------	-----------

In fact, such procedure may cause problems to the students who might be misled by such instructions. Concerning the instructions of the questions, they must be clear, simple, and short.

However, we cannot say that the test did not fulfill its aim-goals. It could test, for example, students' ability to recognize the correct spelling of some words which are usually confusing, focusing on the form. It could also test their abilities to recognize words in groups and classify them according to their meanings. In other words, it could test meaning in all senses of the word.

3. Summary, Conclusion and Recommendation

3.1 Summary

Tests are supposed to treat the issue of what do we want examiners to perform like language for survival needs e.g., eating, direction, hotel, etc. (Valette, 1977). For example, for eating, we need to know something about the literature of feeding, analyze it, etc. There are many types of tests depending on the nature and method of the test. We should also differentiate classroom tests from specialized ones administered to testees in the universities as in the former where there is specific content and the form of the test is also applicable. However, university tests as we have seen are designed to achieve certain goals towards certain goals. Instructors should state their aims in terms of what do they want the testee to know. Is he/she expected to listen to radio station? Is he expected to speak everyday English? etc. Also in terms of skills, sub-skills, linguistic units e.g., prepositions, pronouns, etc., topics e.g., general topics, academic topics, etc., and also in terms of tasks like finding a room in a hotel, managing himself in a restaurant in ordering food, etc.

Again, instructors, as test makers, have to ask themselves: What do we want the testee to do? We give him something that is related to the issue we are going to test him/ her in. However, we create the test i.e. prepare a draft or write it, give him/her a sample, and then we go back to the specifications to see what do we want him/her to do. Unfortunately what happens here is that some instructors violate such rules and principles agreed by most linguists. Before writing (constructing/ preparing the test), a tester should decide on testees and objectives (specific objectives) in terms of what he/she is going to test? When? How? To whom? Why? The second step of course is that he/she has to decide on the format of the test. Testers should realize that it is not easy to make a test; therefore, they should pay a lot of attention, especially if we know that the target is university students and/ or if the test is supposed to test specific area in the language (a skill, a language component, etc.)

3.2 Conclusions

Prior to any further discussion, we should note that the mistakes of these university instructors are committed due to the fact that either they ignore the rules and conditions of language testing or they are unaware of them and that what seems to be the truth. These violations can be described in light of the test purposes. University instructors also did not take into consideration that in testing evaluation, the validity of their tests, relative purpose, reliability, applicability/ practicality, economy, acceptability, relativity, validity, reliability, etc.

will be judged. When they designed their tests, they failed to make them standardized. Standardization means that test should be given out in a large scale, and then the answers should be analyzed with reference to the four phases, the requirements of the good test and norm referenced. In fact, what these testees lack of in designing their tests is the basic studies and / or frame works that they can get from many institutions like American Council on the Teaching Foreign Languages (ACTFL) performance guidelines, Common European Framework, etc.

3.2.1 Receptive Skills' Mistakes

Some instructors think that using short answer questions for testing receptive skills is appropriate, because of their easiness be it when constructing them or be it when marking them. These instructors claim that this type of questions (objective questions) prompt intensive reading; therefore, intensive understanding. What these instructors fail to recognize that these questions do not test students' language; rather, they test their knowledge of the language. Instructors also think that table completion questions are useful for receptive skills as they enhance extensive reading and/ or understanding. Again, they make a mistake as they ignore the fact that in addition to their demerit in involving some writing, students also have to decide on what answers should be written and which will, therefore, be accepted as correct ones.

It is also observed that the instructors under the question made use of pictures thinking that they (pictures) are close to reality and also good for examining specific information. Interdisciplinary, misusing certain information either by testers or even by the testees may lead to converse results. Some instructors ask testees to write in terms of listing, and this is good especially for listening comprehension, but testers here ignore that such questions examine students' ability to recognize words, not their understanding or their meanings.

Some instructors make use of T/F questions due to the fact that they are easy to mark (not to construct of course), but they forget that these question depend upon how high guessing of the testee rather than what language he possesses. For the same reason we find some other instructors using MCQ, but again, these instructors ignore the fact that the wrong distracters they used encourage guessing elements in most of their testees including smart ones. Some other instructors ask testees to listen and complete the text claiming that such questions are good for listening, but again, they forget that these questions examine students' knowledge of language and not the language itself. In other words, they do not test what is supposed to test which make such tests invalid.

We also find that some receptive skills' instructors ask students to identify certain words from amongst many in the text. It is true that this type of questions is good for gist reading in addition to some other features like cohesion, coherence, etc., but they (questions) require thinking which may affect testees' performance because the test will be time-consuming. Additionally, this can be more a test of recognizing words than of understanding them. For those instructors who ask student to link certain items to each other, we need to know what they want testees to perform. Do they want them to prove their ability to intensive reading and/ or understanding? If so, then they (instructors) should number their items rather than

writing them randomly as some of them have done.

3.2.2 Productive Skill's Mistakes

3.2.2.1 Writing

Looking at vocabulary tests, we find that testers seem to be confident of their students' ability to think broadly. In fact, this is one of the demerits of those instructors, because students are beginners and the questions of the instructors seem to be designed for advanced learners. However, such questions are not realistic as they often tests students' imagination or knowledge. The instructors also did make use of guided compositions that are known for the appropriate to beginners. These tests seem to be realistic, but what they do not want to admit is that because of the misuse of instructions, students might not understand the questions, which will lead to wrong output. The instructors also seem to be unaware of other important questions like those on punctuation marks and/ or summary, taking notes questions, dictations (full word and spot), open ended question, integrated questions, etc.

3.2.2.2.1 Speaking

Instructors made use of visual aids mainly pictures to give their students something concrete, but somehow they ignore the fact that this type of tasks is not realistic (means artificial) where testees do not have time to interact with each other. Some other instructors have differently used these visual aids by putting questions based on transferring information through them (pictures, photographs, etc.) thinking that because it is close to reality whereby communication is needed (Carrol, 1980; Weir, 1990). However, what these instructors did not bear in their mind as a demerit of this type of exercises is that it might not be helpful all the time. The reason is so simple: Some students are weaker than others which make such questions worthless, useless, etc.

3.2.3 Language Components' Mistakes

3.2.3.1 Grammar

In order to test grammar, some requirements should be followed. Unfortunately, some instructors misused many of these requirements and/ or criteria. They could not decide, for example, which structure they are testing. It is for this reason perhaps that we miss many rules in the test like if conditional rule, active passive rule, rules of while, when, after, before, etc. We have seen MCQ and testee's own production, but it is not clear whether the former tests student's recognition or whether the latter tests his/her production. It is almost agreed by most linguists that sentence completion usually comes at the end though none of the tests included this type of questions.

Instructors also used questions of mistakes identifications, but they did not inform their students what to do. Do they have to simply identify these mistakes and correct them? They also did not provide questions on transformation, joining sentences or clauses either coordinators (compound sentences) or relative pronouns (complex sentences). Instructor also did not make use of expansion sentences so to test word order which could be, as we know, an adjective (adj.), an adverb (adv.), etc.

Instructors did not make use cloze procedure. It is true that this type of questions is typically used for reading comprehension, but it can be used in grammar (Grammar through cloze procedure) like when we give a list so that students fill in blanks in the passage, MCQ, etc. They also misused the question of matching be it with pictures or vice versa.

3.2.3.2 Sound System

The instructors made use of minimal pairs to test word discrimination or word recognition. The problem here is that they misused them like when they ask students to produce the sound, and we know that a testee cannot produce the sound unless he recognizes it. The instructors also made use of matching oral sound with written word and / or picture, but the mechanism how they conducted that was not clear.

Despite of the importance of finding a match to a key model utterance both for intonation and stress, the instructors did not make use of it. The same thing applies to words that rhyme either those with different intonation patterns or those with different stress patterns. When they wanted to test production, instructors were makes exercises on imitation and repetition, but they did not make it clear to their testees how they want them to do it. Is it by listening and/ or repeating? They should also make use of reading aloud. Some instructors made use of free production and that is a good thing, because this type of exercises can be used both for recognition and production.

3.3 Recommendations

The researcher recommends that testers should consider subjective tests, because we cannot evaluate conversations, dialogues, interviews, etc. Since university tests vary according to the type of the syllabus, instructors should develop a phase of monitoring i.e. to see the effectiveness of their tests on their students, and then analyze the results. They should have band scales and they should give evidence by justifying every single thing they say in the bands. When evaluating the work of the student, students' mistakes should be taken into consideration. Generally speaking, if there are many mistakes (unsatisfactory), we, as instructors, should go back to check out the items to see whether they fulfill our objectives or not.

In a nutshell, the performance of the test is evaluated, not the test itself (Carroll & Hall, 1985). Of course, this evaluation has to be done in light of the testees themselves. In clearer terms, we (as test experts) evaluate the tests in terms of what is wrong with our test characteristics (reliability, validity, discrimination, practicality, accountability, etc.) and also the wash back effect (the influence of the test both on teaching and learning) which can be positive or negative. The goal is to see to what extent these criteria have been functioned. Ultimately, it should be noticed here that more research is needed on how tests can assess what we want students to know, and what we have taught them, and also the impact of test on students' performance.

References

- Acosta, S. (2011). High-stakes reading assessment and English oral language development: A study of third grade English Language Learners in a Texas school district. *Texam A&M Univesity*, 4. Publication No. AAT 3416136).
- Adams, M. Foreign Service (Dept. of State), w.t. (Dept. of State), Washington, DC. Foreign Service Inst. (1978). Testing Kit. *School of Language Studies*, 86.
- Al-Ghamdi, G. (1993). *Language Proficiency Testing: A Historical Approach*. Riyadh: Maktabat al-Tawbah.
- Alonzo, J. (2011). An examination of early reading skill covariance structure invariance across Spanish-speaking English language learners in grades 3 and 4 and their native English-speaking peers. *University of Oregon*, 21. Publication No. AAT 3259271.
- Anivan, S., & Southeast Asian Ministers of Education Organization (Singapore). Regional Language, C. E. (Singapore). (1991). Current Developments in Language Testing. *Anthology Series*, 25.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bolger, D. (2011). The development of orthographic knowledge: A cognitive neuroscience investigation of reading skill. *University of Pittsburgh*, 14-18. Publication No. AAT 3270122.
- Carroll, B. (1980). *Testing Communicative Performance*. Oxford: Program Press.
- Carroll, B., & Hall, P. (1985). *Make Your Own Language Tests: A Practical Guide to Writing Language Performance Tests*. Oxford : Oxford University Press.
- Chapman, A. (2011). Examining the effects of pre-kindergarten enrollment on kindergarten reading readiness. *Tennessee State University*, 13-19. Publication No. AAT 3433392.
- Chen, J. (2011). Language Assessment: Its Development and Future--An Interview with Lyle F. Bachman. *Language Assessment Quarterly*, 277-290.
- Davis, B. (2011). A qualitative study of how writing is used in Catholic secondary schools to foster students' metacognitive skill development. *University of San Francisco*, 21-23. Publication No. AAT 3388382.
- Dennis, D. (2011). Putting the horse before the cart: Utilizing what assessment data reveal about struggling young adolescent readers to inform policy and instruction. *The University of Tennessee*, 8-13. Publication No. AAT 3286915.
- Esmat, G. (2011). "That graphic organizer is awesome" A study examining second grade writers using a graphic organizer and checklist to improve their writing. *University of California*, 13-19. Publication No. AAT 1471218.
- Fergus, D. (2011). Scaffolding instruction and using graphic organizers for writing effective in-class summaries. *University of California*, 3-10. Publication No. AAT 1471212.

- Gautam, A. (2011). Traditional versus Recent Trends in Evaluation relating to Language Teaching. *Language in India*, 62-70.
- Gilfert, S., & Harada, K. (1992). Two Composition Scoring Methods: The Analytic vs. Holistic Method. *Bulletin of Faculty of Foreign Languages*, 17-22.
- Glen, N. (2011). Writing in elementary school science: Factors that influence teacher beliefs and practices. *Syracuse University*, 8-13. Publication No. AAT 3333568.
- Harley, B., & And, O. (1990). National Core French Study: The Evaluation Syllabus. *Canadian Association of Second Language Teachers*, 14.
- Harrison, A. (1980). *A Language Testing Handbook*. London: Macmillan Press.
- Heaten, J. (1988). *Writing English Language Tests*. Harlow: Longman.
- Heckman, J. (2011). *Writing progression of students with limited English proficiency on Texas state exams*. Texas A&M University, 18-22. Publication No. AAT 3405825.
- Heining-Boynton, A. (1990). The Development and Testing of the FLES Program Evaluation Inventory. *Modern Language Journal*, 432-439.
- Hughes, A. (2002). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Inn-Chull. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 39-62. <http://dx.doi.org/10.1177/0265532207083744>.
- Jacobs, H. (1981). *Testing ESL Composition: A Practical Approach*. Rowley: Newbury House Publishers.
- Jang, E. (2005). A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL. *University of Illinois at Urbana*, 33-38. Publication No. AAT 3182288.
- Johansson, S. (1973). An Evaluation of the Noise Test: A Method for Testing Overall Second Language Proficiency by Perception under Masking Noise. *International Journal of Applied Linguistics in Language Teaching*, 78.
- Johansson, S. (1975). Papers in Contrastive Linguistics and Language Testing. *Lund Studies in English*, 50.
- Kohonen, V., & And, O. (1988). Evaluation and Testing in the Learning and Teaching of Languages for Communication. *Education & Culture*, 12.
- Lassche, G. (2006). The Test Usefulness Feedback Framework: Developing and Evaluating a Procedure for Describing Measures of Student Achievement Employed in Web-Based Language Learning Environments. *Online Submission*, 44.
- Lawson, E. (2011). Teacher attitude toward Black English and its impact on reading achievement. *Texas A&M University-Commerce*, 17-26. Publication No. AAT 3445871.

- Lee, L. (2008). Development and validation of a reading-related assessment battery in Malay for the purpose of dyslexia assessment. *Annals of Dyslexia*, 14-17. Document ID: 1512367571.
- Liyang, C. (2008). The key to success: English language testing in China. *Language Testing*, 15-37.
- Madsen, H. (1983). *Techniques in Testing*. Oxford: Oxford University Press.
- Massar, E. (2011). *A case study using the Corrective Reading Program in a junior/senior high remedial class*. Widener University, 7-12. Publication No. AAT 3363759.
- McNamara, T. (1991). The Role of Item Response Theory in Language Test Validation. *ijal*, 44.
- Mercier, D. (2007). The effects of "A Focused Approach to Frontloading English Language Instruction" on the writing skills of fourth grade English language learners. *The Claremont Graduate University*, 19-23. Publication No. AAT 3246656.
- Oller, J. (1983). *Issues in Language Testing Research*. Rowely: Newbury House Publishers.
- Oller, J., & Perkins, K. (1978). *Language in Education: Testing the Tests*. Rowely: Newbury House Publishers.
- Peters, C. (2002). Evaluating cross-language systems the CLEF way. *Cultivate Interactive*, 6. <http://dx.doi.org/10.1016/j.ipm.2007.01.024>.
- Qian, D. (2008). English language assessment in Hong Kong: A survey of practices, developments and issues. *Language Testing*, 85-110. <http://dx.doi.org/10.1177/0265532207083746>.
- Rahman, M. (2011). Making of New Methods for Alternative Assessment and Evaluation in Indian Context: A Perspective. *Language In India*, 217-224.
- Rahman, M., & Gautam, A. (2012). Testing and Evaluation: A Significant Characteristic of Language Learning and Teaching. *Language in India*, 432-442.
- Ramanathan, H. (2008). Testing of English in India: A developing concept. *Language Teasting*, 111-126. <http://dx.doi.org/10.1177/0265532207083747>.
- Raynolds, L. (2011). Differential acquisition of English letter-sound correspondences in bilingual and monolingual primary students. *Fordham University*, 12-17. Publication No. AAT 3262837.
- Roberts Frank, A. (2011). The effect of instruction in orthographic conventions and morphological features on the reading fluency and comprehension skills of high-school freshmen. *University of San Francisco*, 9-12. Publication No. AAT 3326011.
- Robinson, D. (2011). A correlation study between below-grade-level reading and graduation rates in a southeast Georgia school system. *Capella University*, 18-24. Publication No. AAT 3349796.

Sasaki, M. (2008). The 150-year history of English language assessment in Japanese education. *Language Testing*, 63-83. <http://dx.doi.org/10.1177/0265532207083745>.

Secolsky, C., & Dension, D. (2011). *Handbook on Measurement, Assessment, and Evaluation in Higher Education*. Routledge: Taylor & Francis Group.

Simpson, J. (1972). Evaluation and Testing in Modern Languages with Particular Reference to Tests to Accompany the Nuffield/Schools Council Modern Language Materials. *Audio-Visual Language Journal*, 52.

Underhill, N. (2005). *Testing Spoken Language*. Cambridge: Cambridge University Press.

Uysal, H. (2009). Developments in Validity Research in Second Language Performance Testing. *Buckingham Journal Of Language & Linguistics*, 61-68.

Valette, R. (1977). *Modern Language Testing*. New York: HBJ.

Walter, N. (2011). The effects of intervention in phonemic awareness on the reading achievement of English Language Learners in kindergarten. *Northcentral University*, 13-19. Publication No. AAT 3400882.

Weir, C. (1990). *Communicative Language Testing*. London: Prentice-Hall.

West, C. (2011). Exploring interests: Are there relationships among general interests, reading interests, and personality dimensions? *Texas A&M University*, 3-9. Publication No. AAT 3333786.

Glossary

AGI (Arts' Grammar Instructor) = Instructor of grammar at the College of Arts & Linguistics.

ALI (Arts' Listening Instructor) = Instructor of listening at the College of Arts.

ARI (Arts' Reading Instruction) = Instructor of reading at the College of Arts.

ASI (Arts' Speaking Instructor) = Instructor of speaking at the College of Arts & Linguistics.

Assessment= A process in which you make a judgment about a person or situation.

AVI (Arts' Vocabulary Instructor) = Instructor of Vocabulary Building material at the College of Arts & Linguistics.

ERI (Education's Reading Instructor) = Instructor of Reading at the College of Education.

EGI (Education's Grammar Instructor) = Instructor of grammar at the College of Education.

ELI (Education's Listening Instruction) = Instructor of listening at the College of Education.

ESI (Education's Speaking Instructor) = Instructor of Speaking at the College of Arts & Linguistics.

EVI (Education's Vocabulary Instructor) = Instructor of Vocabulary Building material at the

College of Education.

T/F (True or False) = A type of question whereby an examinee decide whether the statement given is true (correct) or false (incorrect). A student may tend sometimes to answer in terms of ticks (✓ for true and □ for false), but this is considered informal.

Copyright Disclaimer

Copyright reserved by the author(s).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).