Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.8, No.30, 2017

Innovative Analytical Assessment: The Relationship Between English E-assessment CBT Scores and Traditional Methods of Assessment

Aadil Askar
Preparatory Year Deanship, King Saud University, Riyadh, Saudi Arabia

Abstract
King Saud University's (KSU) Preparatory Year (PY) examines around 12,000 students per year. The sole objective of the PY is to enable students to progress to the main campus of KSU and complete their degrees; and the main focus of its program is English language acquisition, through the English Language Skills Department (ELSD). Examining such a large number of students every year puts a strain on the resources of the faculty as a whole. As a consequence, maximizing the efficiency of the examination process is of paramount concern. Computer-based testing (CBT) is an extremely effective way of monitoring the knowledge acquisition of a large body of students. Alongside the CBT the PY also implements a speaking and writing examination for the ELSD. This study finds, through a quantitative analysis of statistical data collected from the results of ELSD Humanities students' PY examinations, and multiple regression analysis, a high degree of correlation between the demonstration of proficiencies in the speaking and writing examinations and that of the CBT. This study proposes to measure students' English proficiency through the exclusive use of CBT assessment. Schema theory is utilized to illustrate how the acquisition and organization of knowledge by students can be analyzed effectively. The study takes account of research to date in the field of Instructional Design and Technology to identify the need for learning solution through analysis. The study calls for further research to identify elements of ESL examinations that are overly tedious and inefficient in determining linguistic competencies, using a regression statistical model. It also calls for a more sophisticated statistical analysis, to explore the interrelations in language skills in more detail in order to provide the most rigorous and relevant evaluative framework.
Keywords: Instructional Design, Learning Analytics, Educational Data Mining, Learner Characteristics, e-Assessment.

1. Introduction

In 1957 King Saud bin Abdulaziz Al Saud founded Riyadh University. Eventually renamed King Saud University (KSU) in 1982, KSU developed into an internationally accredited institution. The twenty-first century has ushered in an increasing demand for Saudi citizens to attain higher education qualifications and acquire English in order to remain competitive in the domestic and international marketplace. This period has also seen the foundation of KSU's Preparatory Year (PY), the sole purpose of which is to enable students to progress to the main KSU campus and complete their degree courses. Though the PY implements three primary academic disciplines: English, Self-Development, and Mathematics, special emphasis is afforded to English. The PY offers three tracks to meet the academic needs of its students: Humanities, Medical, and Science. Each track offers a unique curriculum of English instruction to meet the academic expectations of the program each student intends to pursue at KSU's main campus. Most academic fields, such as Science, Engineering, Medicine, and many of the Humanities, are taught exclusively in English at KSU.

KSU's PY offers state of the art infrastructure, qualified global instructors, innovative educational technologies, and a highly professional administration to assist its faculty and students. The PY offers its students a robust higher education, utilizing the most innovative instructional technologies and methodologies. The PY effectively and efficiently trains its native English instructors to use a range of technological hardware and software applications in their classrooms. Interactivity between students and instructors is greatly encouraged in order to promote language acquisition; educational technologies assist in achieving this. The "SMART Board" offers significant interactive audio and visual capabilities, helping the students' uptake of the material presented. For example, an animation of a man walking in a park with his child reframes the language kinetically and a student can define the image for themselves with little instruction. Lexical choices or phrasal variables can be afforded at specific moments of the animation to maximize student learning. Such interactivity presents an alternative model to the traditional, teacher-centered, classroom instruction.

The success of such alternative practices must be determined by periodic testing of the students knowledge. To fulfill this need the PY provides computer based testing (CBT) to measure students' knowledge of English. The CBT tests are limited to vocabulary, reading, grammar, and listening. Speaking and writing are tested separately. These examinations occupy a lot of time in terms of planning and developing the infrastructure required to execute them, and in the time required of the English Instructors to assess and invigilate the examinations. The university overextends its faculty and facilities in order to complete its assessment of its

students, presenting a significant recurring obstacle to the examination of students at KSU.

This study aims to ascertain the usefulness of the CBT, the speaking and writing examinations, and whether they are adequate in assessing students' English proficiency. The study finds a high degree of correlation between the demonstration of proficiencies given by the speaking and writing examinations and that of the CBT. This study aims to show that the absence of writing and speaking examinations display little or no assessment discrepancies. This is demonstrated by the empirical analysis of statistical data collected from the PY examinations. This study proposes to measure students' English proficiency through exclusively utilizing CBT assessment, in order to improve efficiencies. This study uses analytics to provide non-instructional solutions: focusing on the context of the learning process, rather than instructional solutions that concern themselves directly with learning itself.

2. Literature Review

Smith (2011) has argued that assessment is arguably the single most pervasive influence in learning, offering little or no encouragement to advance transference of instruction. For example, as Knight (2002) argues, universities expect their students to strive for a prescribed standard with little or no regard to intellectual competency. Though traditional assessment only measures a fraction of a student's competency, examination grades are the enablers of student progression and are thus the primary focus. The issue with traditional assessment is that the model is heavily determined by the quality of the instructor's input. Milnes and Cheng (2008) assert that research into classroom assessment practices suggests that there is a divergence between approaches recommended by measurement specialists and those actually used by classroom teachers.

Smith (2011) argues that the assumptions that the methodology is a valid or an effective measure of learning is fundamental to the technical and equitable use of summative assessment. Stobart (2005) argues that the case of language is subtler, but a discriminative interaction of examination and English as a second language is a cause for concern. To be effective, assessment must recognize the diversity of learners and allow for differences in styles and rates of learning. To accurately assess English language learners, variations in students' English language skills along with other growth and developmental variations based on their age must be considered. Developmentally appropriate assessment calls for the use of a range of assessment strategies because English language learners need a variety of ways to demonstrate their understanding. The lower the language proficiency, the more important it is to use assessment techniques beyond pencil and paper tasks (Alberta Education 2012).

Huang (2009) argues that when examining English as a Second Language (ESL) writing tasks, the presence of two factors must be taken into consideration: rater related and rask related factors. Rater relation is the confluence of many different considerations, including the rating criteria, the evaluator's academic background, their professional experiences, linguistic background, tolerance for error, perceptions, and expectations. Task related factors include types and difficulty levels of writing tasks (Huang 2009). These factors demonstrate the potential for the evaluators' own bias when grading examinations. Consequently the need for alternative assessment methods must be taken into account.

Examining speaking examinations is more complex. Evaluators utilize a speaking assessment to depart from the tiresome and overemphasized multiple choice exam. An alternative to traditional ESL assessment is translation. Cakir (2013) argues that, when implemented properly, translation displays a genuine measure of a student's performance, through the requirement of the student to replace a text in a source language with a semantically and pragmatically equivalent text in a target language. As such, Bruton (1999) contends, translation is a reconstruction and recontextualization task. With multiple choice examinations an holistic determination of competence is often elusive.

3. Methodology

3.1 Problem statement

Learning environments in higher education demand new approaches in all aspects of learning including innovative assessment. Thus it is important for higher education institutions to implement effective and efficient assessment methods. The KSU PY program conducts at least five CBT exams for around 12000 students every year, with separate speaking and writing exams. This large body of students necessitates the expenditure of large amounts of time and effort on the part of the teaching staff. This study aims to explore an effective, economic, and efficient method of measuring students' English proficiency by focusing on the interrelatedness of English skills. The study examined a large sample of scores from English exams undertaken at the PY. The goal was to reveal any interrelationship between English skills; particularly writing and speaking scores on the one hand and grammar, vocabulary, reading, and listening scores on the other. The findings may ultimately provide evidence to utilize the applicable assessment variables.

3.2 Research Questions

RQ1. Is there a relationship between students' writing scores and their reading, listening, grammar, and vocabulary scores?

RQ2. Is there a relationship between students' speaking scores and their reading, listening, grammar, and vocabulary scores?

RQ3. To what extent and in what manner do the reading, listening, grammar, and vocabulary scores explain the writing scores?

RQ4. To what extent and in what manner do the reading, listening, grammar, and vocabulary scores explain the speaking scores?

3.3 Participants

4138 Humanities track students' scores were selected, with ages ranging from 18 to 21. Although there were students from both genders, the majority of the participants were male. Most of the participants were Saudi nationals living in Riyadh, Saudi Arabia.

3.4 Procedure

The CBT exam was a multiple choice ninety-minute exam focused exclusively on listening, grammar, and vocabulary. It was distributed via an electronic system to examine students over several KSU campuses, including both male and female campuses. The exam was categorized according to the respective tracks the students followed: Humanities, Medical, and Science. In contrast to the CBT the speaking exam comprised an interview with two instructors present. Examiner number one stated questions to solicit responses from the students. Examiner number two had no communicative contact with the students, acting purely as an assessor. The sole purpose of the second examiner was to assess the student's performance and to formulate an overall holistic performance score, in consultation with the first examiner. While the writing exam was based on a single question referencing a commonality of the students' environment and did not have an electronic component.

3.5 Data Collection

Institutional permission has been obtained to conduct the study. The students' scores for the first semester of the 2013-2014 academic year were collected from the PY e-Assessment Center as well as the English department in a digital format. The students' names were excluded from the data to increase the confidentiality of their scores.

3.6 Analysis

A quantitative statistical analysis was conducted, focusing on the correlation and regression between English skills. Correlation analysis was conducted to ascertain the relationship between writing scores and reading, listening, grammar, and vocabulary scores (RQ1). The analysis revealed that there is a weak correlation between writing and the skills assessed in the CBT (Table 1). The highest correlation was between writing and listening, $r = .56$, $p < .001$. While the lowest correlation was between writing and grammar, $r = .468$, $p < .001$.

Table 1. Correlation between writing, reading, listening, grammar and vocabulary scores.

Writing correlation matrix

	Writing	Reading	Listening	Grammar
Reading	0.549			
	0.000			
	0.560			
Listening	0.719			
	0.000			
	0.000			
Grammar	0.468	0.602	0.551	
	0.000	0.000	0.000	
Vocabulary	0.490	0.737	0.663	0.634
	0.000	0.000	0.000	0.000

Cell Contents: Pearson correlation
P-Value

$p < 0.001$.

Table 3. Regression analysis

Model	Variables Entered	Variables removed	Method
1	Listening	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Reading	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	Grammar	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

To ascertain the relationship between the speaking scores and the reading, listening, grammar, and vocabulary scores, a correlation analysis was performed. The analysis indicated that there is a weak correlation between the speaking scores and the reading, listening, grammar, and vocabulary scores (Table 2).

The analysis indicated that listening scores explains 31.3% of the variation in the writing results, the listening and reading scores explain 35.8% of the variation in the writing scores, and the listening, reading and grammar scores explain 37.4% of the variation in the writing scores (Table 4).

Table 4. Variation of test scores

Model Summary

					Change statistics				
Model		Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	F1	F2	Sig. F Change
1	.560[a]	.313	.313	1.42367	.313	1886.921		4136	.000
2	.598[b]	.358	.358	1.37678	.045	287.538		4135	.000
3	.611[c]	.374	.373	1.36012	.016	102.870		4134	.000

a. Predictors: (Constant), Listening
b. Predictors: (Constant), Listening, Reading
c. Predictors: (Constant), Listening, Reading, Grammar

Multiple regression analysis was employed to examine to what extent and in what manner the reading, listening, grammar, and vocabulary scores explained the speaking scores (Table 5). The statistical analysis revealed that all independent variables contribute to the speaking scores

Table 5. Multiple regression analysis

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Listening		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Reading		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	Grammar		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	Vocabulary		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Speaking

The analysis showed that the listening scores explain 34.0% of the variation in the writing results, the listening and reading scores explain 37.9% of the variation in the writing scores, and the listening, reading and grammar scores explain 39.4% of the variation in the speaking results. The four independent variables listening, reading, grammar and vocabulary explain 40.1% of the variation in the speaking results (Table 6).

Table 6. Variation of test scores.
Model Summary

Model		R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.583[a]	.340	.340	2.86199	.340	2134.625	1	4136	.000
2	.616[b]	.379	.379	2.77645	.039	259.769	1	4135	.000
3	.628[c]	.394	.394	2.74378	.015	100.043	1	4134	.000
4	.633[d]	.401	.401	2.72767	.007	49.980	1	4133	.000

a. Predictors: (Constant), Listening
b. Predictors: (Constant), Listening, Reading
c. Predictors: (Constant), Listening, Reading, Grammar
d. Predictors: (Constant), Listening, Reading, Grammar, Vocabulary

Where we note the following:
1- The independent variable listening explains 31.3% of the variation of writing results only, which is highly significant, as indicated by the value of statistics F.
2- The independent variables listening and reading explain 35.8% of the variation of Writing results only, which is highly significant, as indicated by the value of statistics F.
3- The independent variables listening, reading and grammar explain 37.4% of the variance results of writing only. Meaning that 62.6% of the variation in writing results support the contrast capabilities of individual students which is highly significant as it refers statistic value F.
Table 2. Correlation between speaking, reading, listening and grammar scores.
Speaking correlation matrix

	Speaking	Reading	Listening	Grammar
Reading	0.569			
	0.000			
Listening	0.583			
	0.753			
	0.000	0.000		
Grammar	0.530	0.691	0.671	
	0.000	0.000	0.000	
Vocabulary	0.548	0.732	0.715	0.679
	0.000	0.000	0.000	0.000

Cell Contents: Pearson correlation
P-Value
$p < 0.001$.

In order to investigate to what extent and in what manner the reading, listening, grammar, and vocabulary scores explain the writing scores, a multiple regression analysis was conducted (Table 3). The analysis showed that three independent variables (listening, reading and grammar) explain the writing scores.

3.7 Limitations
The Humanities, Science, and Medical tracks are examined simultaneously. This presents difficulty in assessing the students overall performance collectively, since each track follows a different curriculum. The statistical data parses out distinctive variables to present a more genuine measurement of KSU students' performance.

4. Conclusion
Saudi Arabia is committed to integrating twenty-first century learning systems into its educational framework. As more students enter higher education, the need to implement models that can determine the effectiveness of the learning taking place, in order to make improvements and adjustments where necessary, is of paramount importance. Finding a solution that can assess the students' performances quickly, efficiently and accurately is crucial to maintaining rigorous academic standards. On a more local level, within the remit of this study, the need for such streamlining in testing is emphasized by the overstretching of the ESL faculty at the PY during examinations, especially with regards to facilitating the cumbersome speaking and writing exams.

This research attempted to establish whether there is a relationship between students' writing scores and their reading, listening, grammar, and vocabulary scores. There was a need to ascertain the importance of the speaking and writing examinations, specifically their effectiveness in establishing the knowledge acquired. Schema theory provides a model through which one can analyze how knowledge is acquired, processed and

organized. This study goes some way to addressing this need through the use of schema theory.

This research identifies KSU's successful implementation of a regression statistical model, focusing on vocabulary, grammar, reading, and listening. Although the research model concludes that there is a significant relation between the variables, a further study is needed to further explore the interrelations in language skills to achieve more accurate explanations. Specifically, a more sophisticated statistical analysis is needed. Though this study has demonstrated a significant relationship between the variables outlined, other methods of statistical analysis could help explain why these relationships exist.

This researcher calls on further research, using the regressive modal to identify elements of ESL examinations that are overly tedious and inefficient in determining linguistic competencies. Furthermore, though significant correlations may exist between CBT, speaking and writing examination scores, this should not indicate the lack of any need to assessment of the speaking and writing components. Rather different methods for assessment should be considered in order to provide the most rigorous and relevant evaluative framework.

References

Bruton, A. (1999). Comments on James D. Brown and Thom Hudson's "the Alternatives in language assessment" a reader reacts. TESOL Quareterly, 33 (4), 729-734.

Çakir, Cemal. "Standard Assessment And Alternative Assessment In English Language Teaching Program." *Gazi University Journal Of Gazi Educational Faculty (GUJGEF)* 33.3 (2013): 531-548. *Education Research Complete*. Web. 26 Sept. 2014.

Cheng, Liying, and Terry Milnes. "Teachers' Assessment Of ESL Students In Mainstream Classes: Challenges, Strategies, And Decision-Making." *TESL Canada Journal* 25.2 (2008): 49-65. *ERIC*. Web. 26 Sept. 2014.

Jinyan, Huang. "Factors Affecting The Assessment Of ESL Students' Writing." *International Journal Of Applied Educational Studies* 5.1 (2009): 1-17. *Education Research Complete*. Web. 26 Sept. 2014.

Knight, PT. "Summative Assessment In Higher Education: Practices In Disarray." *Studies In Higher Education* 27.3 (n.d.): 275-286. *Social Sciences Citation Index*. Web. 26 Sept. 2014.

Moores-Abdool, Whitney, Noorchaya Yahya, and Caridad H. Unzueta. "Learning Preferences of Saudi University Students With Native English Speaking Teachers. (Cover Story)." *Perspectives (TESOL Arabia)* (2011): 18-24. *Education Research Complete*. Web. 26 Sept. 2014.

None. *Language Proficiency Assessment*. N.p.: None, n.d. *Supporting English Language Learners*. Alberta Education, 2012. Web. 25 Sept. 2014. <http://learnalberta.ca/content/eslapb/languageproficiencyassessment.html>.

Smith, Chris. "Examinations And The ESL Student--More Evidence Of Particular Disadvantages." *Assessment & Evaluation In Higher Education* 36.1 (2011): 13-25.

ERIC. Web. 26 Sept. 2014.

Stobart, Gordon. "Fairness In Multicultural Assessment Systems." *Assessment In Education: Principles, Policy And Practice* 12.3 (2005): 275-287. *ERIC*. Web. 26 Sept. 2014.