Comparing Predictive Validity of Islamic Azad University English Proficiency Test and Standard Proficiency Tests against a Socio-Cognitively Validated Test of English for Specific Purpose

Khalil Motallebzadeh[1] and Mahmood Khosravani*[2]

[1]Associate Professor, Department of English, Tabaran Institute of Higher Education, Mashhad, Iran
[2]PhD candidate, Department of English, Torbat-e Heydarieh Branch, Islamic Azad University, Torbat-e Heydarieh, Iran
Corresponding Author: Mahmood Khosravani, E-mail: sadradone@gmail.com

ARTICLE INFORMATION	ABSTRACT
	In spite of the facts that validity is known as a unified concept in contemporary theory and practice, investigating the sub-construct of this concept enlightens the discipline to a great extent. Predictive validity of high-stakes tests such as English proficiency tests (EPTs) which are used for critical educational decisions is an example. EPT of Islamic Azad University (IAU), as a localized form of national and international scale proficiency tests, is served as an obligatory part of exit program for PhD candidates. Considering the fact that manipulating ESP texts is a major language learning goal in the university, the current research was an attempt to examine the predictive validity of IAU-EPT against a valid third measure in comparison with other EPT exams. To this aim, adopting Weir's socio-cognitive framework, an ESP test of politics was developed and validated. Then, the test was administered to 19 PhD candidates of politics with IAU-EPT certification and 16 PhD candidates certified at other EPTs (TOEFL, GRE, and Tolimo). Pearson correlation was employed to investigate the relationship between their authorized EPT scores and their performance in a real context validated ESP test. The findings revealed a weak predicting value for IAU-EPT in a ESP task while a strong and significant correlation was reported between standard EPTs scores and ESP test. Further investigation is required to investigate the construct validity of IAU-EPT. Future researchers are recommended to include more participants from other disciplines. These findings have implications for language learners, language teachers, researchers and educational policy makers.

1. Introduction

As soon as second language became a tool for the purpose of instructions of different disciplines of sciences and appeared as a criterion for selection in different academic settings in national and international contexts, second language proficiency and the way it is measured attracted practitioners and researchers (Spearman, 1927; Lado, 1961; Oller, 1983 Canale & Swain, 1980; Fulcher & Davidson 2007; Bachman, 1990; Bachman & Palmer 2010 and etc). Similarly, considering the crucial decisions made on these measures, reliability and validity of these language proficiency tests took on an added importance. According to Chapelle and Voss (2014) a vast bulk of theoretical and experimental studies are conducted on validity and validation of different kinds of high-stakes test in the literature amongst starting from Messick (1989), Kane (1992), Bachman and Palmer (1996), Bachman (2005) to Kane (2013) which inquired into different dimensions of the concept including content, criterion-referenced and construct validities, test-usefulness as well as approaches toward submitting proofs on validity, such as evidence-based approach and argument-based approach. Though the unitary view of validity was theoretically accepted by a large number of scholars in the field, applied researchers carried out several researches investigating the sub-constructs of validity such as content validity, face validity and etc because each of them could bring about valuable information concerning the worth and feasibility of a given test. Among these sub-constructs, predictive validity which has greatest contribution to appropriateness of test use is investigated in too many studies.

In this regard, Hughes (2003) states that for a test to be valid, it is required to measure the thing it purports to measure. In addition, emphasizing that a valid test must estimate and predict the possibility of the examinees' future Brown (2004) asserted that predictive validity of an assessment assumes critical importance in placement tests, admissions assessment batteries, language aptitude tests, and those with similar application. Besides, Fulcher and Davidson (2007) gave greater weight to the predictive validity of high-stakes tests and postulated that validity evidence is the strength of the predictive relationship between the test score and that performance on the criterion.

English proficiency test of Islamic Azad University (IAU-EPT) is an obligatory part of educational program for graduation of PhD Candidates in different disciplines. This test is a nativized version of widespread paper and pencil general proficiency tests which has undergone basic modifications to include only knowledge of vocabulary, grammar and reading comprehension. Elder and O'Loughlin (2003) believe that that research into appropriate test use and the meanings attributed to it seems necessary. Also, Woodrow (2006) states that it is important to investigate the predictive validity of famous proficiency tests such as IELTS to see whether they are predictive of academic performance in specific academic settings. Considering these issues on one side and questioned construct validity of IAU-EPT (due to fundamental modifications) on the other, the current research is an attempt to investigate its predictive validity in comparison with other standard proficiency tests (Tolimo, GRE, TOEFL) against a third measure ESP test developed on the basis of Weirs' socio-cognitive model (2005). To this aim, the following research questions are formulated:

1. Is there any relationship between IAU-EPT scores and the developed ESP test?
2. Is there any relationship between standard proficiency tests scores and the developed ESP test?

2. Literature Review

the emergence of predictive validity dates back to Nuttall (1986) who provided a theoretical explanation of the concept by asserting that for a test to be confirmed as valid, it is required to reasonably predict future academic performance, i.e. a sample of examines' performance in a test must be generalizable to a universe of their understanding in order to assess expected learning. Such an understanding assists examiners to make appropriate decisions and also informs teaching to modify the instruction aimed promoting learning. Messick (1989) brought the necessity of predictive validity into the core of validity research and stated that with respect to the generality of the process, the development of evidence to support an inferential leap from an observed consistency to a construct or theory that accounts for that consistency is a generic concern of all science. This postulation was echoed by Bachman (1990) who asserted that predictive validity resides in heart of construct validity because construct validity concerns the extent to which performance is consistent with predictions that we make on the basis of a theory of abilities, or constructs.

Perceiving the importance of this sub-construct of validity, theorists, researchers and practitioners developed theories, formulated methods of measurement and conducted research from different angles of look. A huge bulk of these attempts was devoted to predictive validity of academic proficiency and achievement instruments amongst which were the study of Cho and Bridgeman (2012) on TOEFL and GPA, study of the predictive validity of GRE by Ortega and Payne (2007), study of IELTS as a predictor of academic performance by Bayliss and Ingram (2007), study of IELTS as a predictor of academic performance by Paul (2007) and Sternberg's study of the predictive validity of the American SAT (2006). In addition, Banerjee (2003) investigated the use and interpretation of proficiency test scores and found that the examinees' initial language proficiency is a good predictive of their study experience. Finally, Feast (2002) studied the impact of IELTS scores on academic performance in university and found a moderate positive relationship.

In Iranian context, however, Mozaffarzadeh, Pourgasem and Gerami (2019) examined the predictive validity of IELTS and performance of Iranian examinees in real context. Findings showed that this test is not an appropriate predictor of successful performance. Also, Alavi (2012) inquired into the predictive validity of final English exams as a measure of success in Iranian national university entrance exam and found was a positive relationship between each of the exams and Iranian national university entrance English exam. Furthermore, Mohammadi (2009) probed the correlation between English Language students' academic achievement regarding their Grade point Average (GPA) and their entrance admission test results and their sex, field of study, or age. The results showed that the students' total score in the examination was a better predictor than the English language subtest which was beyond the researcher's expectation. It was also revealed that GPA of freshmen was also equally predictable by the entrance examination and its English proficiency subtest. Finally, in a similar study conducted by Maleki and Zangani (2007), the findings of the study indicated a positive moderate correlation between performance in TOEFL and GPA of Iranian learners.

3. Method

Considering the fact that English learning in Islamic Azad university is mainly aimed at enabling learners to manipulate (comprehending, translating) texts of English for specific Purpose (ESP), the intent of the current research was investigating the extent to which EPT and some other standardized proficiency tests can predict the linguistic performance of PhD candidates in a real ESP context. To this purpose, a quantitative correlation design was utilized to estimate the predictive validity of EPT in a socio-cognitively developed ESP test.

3.1 Participants

Non-random purposive sampling was adopted for selecting participants. Since those PhD candidates of IAU that are already certified in a standard national or international English proficiency test become exempted from EPT, we selected our participants from two populations: those who had an IAU-EPT authorized score and those who were already certified in another English proficiency test. Accordingly, 16 exempted PhD candidates which were certified in Tolimo, GRE, TOEFL were selected purposively. Also, 19 other PhD candidates who were qualified in IAU-EPT (scored beyond 50) during a year were selected purposively. All these participants were male or female PhD candidates selected form different disciplines of politics in different branches of Islamic Azad University in the city of Tehran. They were in the third year, fourth year or fifth year and more of their educational program.

A) Weir's socio-cognitive framework for test validity: The theoretical framework selected for developing a valid third measure of ESP adopted from Weir (2005) which investigates the examinees performance in a real academic setting. This framework enables language test designers to evaluate the quality of all four language skills incorporated in a test. The framework encompasses five types of validity evidence including context validity, theory-based validity, scoring validity, consequential validity, and criterion-related validity. The ability that needs to be tested is determined by the internal mental process of the test taker and the use of language in the task is presented as a social rather than the only linguistic phenomenon, providing a better framework to consider the consequential validity of a test.

B) ESP Test: a researcher-developed ESP test was used to examine the performance of the participants of the study in a real academic context. The content of the test was selected from major sources of the politics taught in different branches of IAU. The face validity and content validity of the test was verified by a panel of experts in the university. Besides, the reliability of the test was estimated in a pilot study on 34 M.A and PhD candidates of different disciplines of politics in IAU (Cronbach Alpha=0.77). More importantly, the test validated against the adopted framework (Weir, 2005).

3.2 Procedure

19 participants with IAU-EPT certification and 16 participants with Tolimo, GRE and TOEFL were selected among PhD candidates from different disciplines of politics in different branches of IAU in Tehran. These students were administered a validated ESP test of politics and their performance in the test was investigated. The scores obtained in the ESP test served as a third measure against which the predictive validity of IAU-EPT and other standardized English proficiency tests were measured. The obtained data were entered into SPSS (version 21) and analyzed through Pearson Product Moment Correlation test.

4. Results and Discussion

This research intended to investigate the existence of any significant relationship between IAU-EPT and other standard tests of proficiency against a socio-cognitively validated ESP test. A Pearson's correlation was run to draw a line of best fit through the scores obtained from the scales. The findings are reported in the following table:

Table 1: *Descriptive statistics for IAU-EPT/ESP correlation*

	Mean	Std. Deviation	N
IAU_EPT	58.47	7.633	19
ESP	11.11	2.720	19

As it is revealed in this table, the mean score for IAU-EPT is 58.47 showing that the participants scores revolve around the cut-score for pass or fail in the proficiency test. In addition, the mean score for their performance in the ESP test was 11.11 which is under the cut-score for pass and fail in PhD courses in the university. This simply indicates that the qualification the IAU-

EPT is not a good predictor for academic performance. Yet, to affirm this conclusion through a robust statistical procedure, the Pearson-Correlation output reported in table 2 is interpreted.

Table2: *Pearson output for IAU-EPT/ESP correlation*

		IAU_EPT	ESP
IAU_EPT	Pearson Correlation	1	.403
	Sig. (2-tailed)		.097
	N	19	19
ESP	Pearson Correlation	.403	1
	Sig. (2-tailed)	.097	
	N	19	19

As it is indicated in this table, the Pearson correlation coefficient, r, is 0.403. Also, as it is represented, this relationship is not statistically significant (p = 0.097). In sum, a Pearson product-moment correlation was run to determine the relationship between IAU_EPT and ESP scores. There was not a strong, positive correlation between these two variables, which was not statistically significant (r = .403, n = 18, p = .097). Also, the relationship between other_EPTs and ESP test was investigated the results of which are reported below:

Table 3: *Descriptive statistics for other-EPTs/ESP correlation*

	Mean	**Std. Deviation**	**N**
Other_EPTs	61.95	8.270	19
ESP	13.17	2.036	19

In table 3, unlike the results represented in table 1, though the mean score of the proficiency is 61.92 which revolve around that of IAU-EPT group, the ESP scores exceed the cut-score for pass and fail in the university.

Table 4: *Pearson output for other-EPTs/ESP correlation*

		Other_EPTs	ESP
Other_EPTs	Pearson Correlation	1	.931[**]
	Sig. (2-tailed)		.000
	N	19	18
ESP	Pearson Correlation	.931[**]	1
	Sig. (2-tailed)	.000	
	N	18	18

The values in table 4 represent a very strong positive correlation between the scores (r = 0.931) which is near to perfect. In sum, a Pearson product-moment correlation was run to determine the relationship between standard EPT scores and a validated ESP scores. There was a strong, positive correlation between these two sets of data, which was statistically significant (r = .931, n = 19, p = .000).

To investigate the predictive validity of standardized English proficiency tests which serve as proficiency criteria in a global scale, many quantitative and qualitative studies are conducted. Some of these studies have reported a strong positive relationship between performance on these tests and performance a real-context task, as it was the case in the present study. However, some researchers conducted in this filed submitted contradictory results amongst which are Dooey and Oliver (2002), Woodrow (2006) and Komba (2012) who reported weak evidences on the existence of this relationship. Similarly, in our study, the nativized from of these tests (IAU-EPT) could not predict the performance in real context while their national and international counterparts reported a strong correlation. It seems that breaking down an English proficiency test into the knowledge of vocabulary, grammar and reading comprehension has undermined the construct validity of the test. Therefore, further studies are required to probe the construct validity of the test n future. Besides, in spite of accepting IAU-EPT not as a proficiency test (in sharp contrast with the brand), this tool, even as tool of English sufficiency,

does not correlate with the academic performance, at least in this little scale study. Modification and policy makings concerning this test need further research.

These findings in the study stand in sharp contrast with Mozaffarzadeh, et al (2019) that concluded a meager predictive validity for IELTS for performance in the real context of language use. Yet, these findings are in line with Rumsey (2013) which implies that administration of localized proficiency tests as placement tool or other high-stakes decision such as admissions must be done with ultimate care because these tests might suffer from robust predictive validity.

5. Conclusion

In an innovative attempt, the current research compared predictive validity of IAU-EPT and other standard proficiency tests against a socio-cognitively validated ESP tests which served as the third measure. The findings showed that performance in IAU-EPT do not predict performance in a real academic task while scores obtained in other EPT tests were highly correlated with those on the task. More number of participants through other sampling procedures is required for generalizability of the research. Further research is suggested to be conducted to delve into the construct validity of localized test (IAU-EPT). These findings have implications for students of humanities, applied linguists, researchers and educational programmers.

References

[1]   Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
[2]   Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, 2*, 1–34.
[3]   Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: designing and developing useful language tests*. Oxford University Press.
[4]   Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford University Press.
[5]   Bayliss, A., & Ingram, D. E. (2006). *IELTS as predictor of academic language performance*. Australian International Education Conference, (pp. 1-12). Retrieved from www.idp.com/aiec.
[6]   Canale, M. and Swain, M. 1980: Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics 1*, 1–47.
[7]   Cho, Y& Bridgeman, B. (2012). Relationship of TOEFL iBT(R) scores to academic performance: Some evidence from American universities. *Language Testing, 29*. 421-442.
[8]   Brown, H. D. (2004). Language assessment: principles and classroom practices. New York: Pearson/Longman.
[9]   Chapelle, C. A., & Voss, E. (2014). *Evaluation of language tests through validation research*. In A. J. Kunnan (Ed.), The companion to language assessment (pp. 1079–1097). Chichester: Wiley.
[10] Dooey, P., & Oliver, R. (2002). An investigation into the predictive validity of the IELTS test as an indicator of future academic success. *Prospect, 17*(1), 36-54.
[11] Elder, c., & O'Loughlin, K. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS. *International English Language Testing System (IELTS): Research Reports, 3,* 207-254.
[12] Feast, V. (2002). The impact of IELTS scores on performance at University. *International Education Journal, 3*(4), 70-85.
[13] Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: an advanced resource book.* Routledge.
[14] Hughes, A. (2003). *Testing for language teachers.* Cambridge University Press.
[15] Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527–535.
[16] Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73.
[17] Komba, S. (2012). Investigation of the predictive validity of communication skills examination on university students' overall academic performance in Tanzania. *International Journal of Education, 4,* 247-266.
[18] Lado, R. (1961). *Language Testing*. McGraw-Hill.
[19] Maleki, A., & Zanjani, E. (2007). A survey on the relationship between English Language proficiency and the academic achievement of Iranian EFL students. *Asian EFL Journal, 9*(1), 86-96.
[20] Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(5), 5-11.
[21] Mohammadi, M. (2009). The Predictive Validity of Islamic Azad. *The Journal of Modern Thoughts in Education, 4*, 59-72.
[22] Mozafarzadeh, S.; Pourgasem, A., & Gerami, E. (2019). *The predictive validity of IELTS and Iranian testees' performance in real context.* IELTS for teachers, Amirkabir University of technology. Iran.
[23] Nuttal, D. (1986). Assessing educational achievement. *British journal of sociology of education, 8*(3), 343-348.
[24] Oller, J. W. (Ed.). (1983). *Issues in Language Testing Research*. Rowely, MA: Newbury House.
[25] Paul, A. (2007). IELTS as a predictor of academic language performance - part 2. IELTS Research Notes. 7.4. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report3.ashx.
[26] Rumsey, L.W. (2013). Shedding light on the predictive validity of English proficiency tests in predicting academic success [Unpublished PhD dissertation]. University of Newcastle, Newcastle.
[27] Spearman, C.E. (1927). *The Abilities of Man: Their Nature and Measurement.* Macmillan.
[28] Weir, C. (2005). Language testing and validation. An evidence-based approach. Palgrave MacMillan.
[29] Woodrow, L. (2006). Academic success of international postgraduate education students and the role of English. *University of Sydney Papers in TESOL, 1*, 51-70.