

ĐÁNH GIÁ BÀI KIỂM TRA TIẾNG ANH CHƯƠNG TRÌNH TIỀN TIẾN TẠI MỘT TRƯỜNG ĐẠI HỌC CÔNG LẬP Ở VIỆT NAM

Nông Thị Hiền Hương, Soubakeavathi Rethinasamy***

Kiểm tra đánh giá là một phần quan trọng của việc dạy và học. Trong lĩnh vực giảng dạy ngôn ngữ, nhiều bài thi tiếng Anh thương mại luôn có sẵn nhưng tương đối đắt và không phù hợp cho từng nhu cầu cụ thể. Do đó, nhiều trường đại học đã xây dựng các bài kiểm tra tiếng Anh cấp cơ sở. Bài nghiên cứu đánh giá ba lĩnh vực: giá trị tiêu chuẩn đồng hành, giá trị dự đoán và giá trị nội dung của bài kiểm tra tiếng Anh chương trình tiên tiến tại một trường đại học công lập ở Việt Nam. Kết quả cho thấy bài kiểm tra tiếng Anh có mối quan hệ tương quan với điểm IELTS và điểm trung bình học tập toàn khóa. Tuy nhiên, giá trị nội dung và phương thức chuẩn bị cho bài kiểm tra ở mức độ trung bình. Các tác giả sẽ thảo luận các bước để nâng cao độ giá trị của bài kiểm tra tiếng Anh. Hy vọng rằng nghiên cứu này được coi là một mô hình đánh giá các bài kiểm tra ngôn ngữ cấp cơ sở.

Từ khoá: Kiểm tra ngôn ngữ, độ giá trị của bài kiểm tra, kiểm tra độ giá trị, xác nhận kiểm tra.

Testing and assessment plays an integral role in teaching and learning. In language teaching, despite their ready availability, many Commercial English proficiency tests seem rather costly and not appropriate for specific needs. Thus, many universities have designed their own English proficiency tests. This study evaluated three types of validity of the Advanced Educational Program English Test (AEPET) at a public university in Vietnam: concurrent, predictive and content validity. The results revealed that AEPET scores significantly correlate with IELTS scores and CGPA; whereas, the content validity and preparation for the test remain moderate. The paper will discuss the steps to further improve AEPET's validity. It is hoped that this research will serve as a model for the evaluation of in-house language tests.

Key words: Language testing, test validity, test validation.

* ThS., Trường Đại học Nông Lâm Thái Nguyên

Email: hhuong04052002@yahoo.com

** PGS., Trường Đại học Malaysia Sarawak (Malaysia)

INVESTIGATING THE VALIDITY OF THE ADVANCED EDUCATIONAL PROGRAM ENGLISH TEST AT A PUBLIC UNIVERSITY IN VIETNAM

1. Introduction

Testing has immense effect on teaching and learning. Thus, if designed and executed properly, tests can help to bring about positive changes to teaching and learning. Weir (2004) states that test validation is the “process of generating evidence to support the well-foundedness of inferences concerning trait from test scores, i.e., essentially, testing should be concerned with evidence-based validity.”(p.2). Therefore, test validation plays the most important role in test development and use and should be always examined (Bachman & Palmer, 1996). In the light of the importance of test validation, this study is aimed to validate an in-house English test conducted at a public university in Vietnam. The AEPET is an English achievement test and frequently carries out at the end of the English language course. In order to investigate the validity of the AEPET, the study intends to assess English language lecturers’ judgments about the AEPET whether the AEPET content reflects the knowledge and skills required to do in the Advanced Educational Program (AEP) syllabus as well as to find out to what extent AEPET’s preparation is present and applicable before the examination is administered. Furthermore, the study also aims to examine the extent to which the

AEPET correlates with International English Language Testing Services (IELTS) scores as well as address the question to what extent the test validation determines academic success for the AEP students at a public university in Vietnam.

The study aims to determine the validity of the AEPET at a public university in terms of the concurrent validity, predictive validity and content validity. The study intends to answer following research questions:

- (1) What is the relationship between the students’ AEPET and the IELTS scores (Concurrent Validity)?
- (2) What is the relationship between the students’ AEPET scores and academic achievement, in comparison with (Predictive Validity)?
- (3) What is the content validity of the AEPET?

3. Methodology

The AEPET consists of four components: Listening, Reading, Writing and Speaking. Each question in each component, in this study, was analyzed by using quantitative methods.

4. Results Interpretation

4.1. Results on Concurrent Validity

The scores from 103 students’ AEPET and IELTS academic transcripts were

keyed into Statistical Software for Social Science (SPSS) version 23 and then descriptive statistics were computed in order to see how the students have performed on each component and overall score in the two tests.

Table 1 Correlation Results between AEPET and IELTS

Component	AEPET vs. IELTS	
	R - value	P - value
Listening	Weak (r=.261)	Significant (p=.008)
Reading	Weak (r=.351)	Significant (p=.000)
Writing	Weak (r=.307)	Significant (p=.002)
Speaking	Moderate (r=.517)	Significant (p=.000)
Overall	Weak (r=.398)	Significant (p=.000)

Table 1 indicates that the relationship between AEPET and IELTS is significant. The strength of significance (P-value) is less than 0.01 level, showing very high significant correlation between each component: Listening, Reading, Writing, Speaking and overall band scores between the two tests. In addition, the correlation coefficient (R-value) which ranges from .261 to .517, is positive, indicating from weak to moderate correlations. More specifically, the highest correlation is found for Speaking Component (r=.517) which is a moderate correlation, followed by Overall (r= 0.398), Reading (r=0.351), Writing (r= 0.307) and Listening (r=.0.261) which are weaker correlations.

4.2. Results on Predictive Validity

The scores of the AEPET, IELTS and CGPA were coded and processed by using SPSS version 23. Firstly, the descriptive statistics were computed in order to see how the students have performed on each component and overall scores as well as their CGPA scores. Secondly, Pearson Correlation was used to determine the correlation between the AEPET and CGPA; between the IELTS and CGPA. Finally, Linear Regression was used to determine the impact of the AEPET and IELTS on students' CGPA. Table 2 presents the predictive validity results for AEPET and IELTS scores.

Table 2 Predictive Validity Results for AEPET and IELTS Scores

AEPET vs. CGPA					IELTS vs. CGPA			
Component	P	R	R ²		Component	P	R	R ²
Whole sample (N=143)	Overall	.000	.613	.376	Overall	.000	.614	.376
	Speaking	.000	.549	.301	Speaking	.000	.535	.286
	Reading	.000	.451	.204	Listening	.000	.471	.225
	Listening	.000	.414	.171	Reading	.000	.428	.183
	Writing	.000	.266	.134	Writing	.000	.314	.099

Table 2 shows that both AEPET and IELTS components significantly correlate with CGPA. The P-value is less than 0.01, showing a very high significant level between each component: Overall, Speaking, Reading, Listening and Writing scores in the two tests and CGPA. In addition, R-values in the two tests are all positive, indicating weak to strong correlations. Across the two tests, strong strength of correlation is found for the relationship between AEPET Overall scores and CGPA ($r=.613$); IELTS

Overall scores and CGPA ($r=.614$), predicting 37.6% of the variance of success in CGPA. Thus, overall scores of the two tests emerge as the most significant predictors for academic success. Likewise, the moderate correlations are observed for the association between Speaking, Reading and Listening scores with CGPA in the two tests. However, weak correlation is accounted for the relationship between Writing scores and CGPA.

Table 3 Regression Results between AEPET scores and CGPA

Component	AEPET vs. CGPA					IELTS vs. CGPA				
	R ²	β	t	F	P	R ²	β	t	F	P
Overall	.376	.613	9.212	84.868	.000**	.376	.614	9.228	85.110	.000**
Speaking	.301	.549	7.795	60.770	.000**	.286	.535	7.513	56.446	.000**
Reading	.171	.414	5.402	29.178	.000**	.183	.428	5.622	31.602	.000**
Listening	.204	.451	6.006	36.073	.000**	.225	.474	6.390	40.828	.000**
Writing	.134	.366	4.674	21.857	.000**	.099	.314	3.932	15.461	.000**

Notes

** : Significant at the 0.01 level (2-tailed)
 Predictors : AEPET Listening, Reading, Writing, Speaking and Overall scores
 Dependent variable : CGPA

Table 3 shows that for both AEPET and IELTS, there is significant correlation between AEPET scores and CGPA; IELTS scores and CGPA because the P-value is less than 0.01, showing a strong regression line between each component, overall band scores of the two tests and CGPA.

Across the two test, the highest coefficient of determination (R-squared) is observed for the agreement between Overall scores and CGPA and then

followed by Speaking scores, indicating approximately from 28 to 37% of the variance of success in CGPA. By contrast, lower R-squared values are found for the relationships between CGPA and Listening scores; Reading scores; Writing scores, corresponding from 10% to 20% of the variance of academic success.

In short, both AEPET and IELTS components significantly correlate with CGPA. Across the two tests, only strong correlation is found for the relationship

between AEPET Overall and CGPA; IELTS Overall scores and CGPA, thus overall scores of the two tests emerge as the most significant predictors for academic success. Likewise, the moderate correlations are observed for the association between Speaking, Reading and Listening scores with CGPA in the two tests. However, weak correlation is accounted for the relationship between Writing and CGPA. Therefore, in a nutshell, it can be concluded that just like IELTS, the AEPET is considered a significant predictor for students' academic achievement.

4.3. Content Validity

The content validity of the AEPET was investigated with a focus of two major two parts: content validity of AEPET components: Listening, Reading, Writing and Speaking and content validity of the process of AEPET's preparation. The results on content validity of AEPET components and AEPET preparation are presented individually as follows.

4.3.1 Content validity for AEPET components

Table 4 Mean Scores for AEPET Components

Component	Overall mean	SD	Degree
Speaking Test	3.56	.860	H
Reading Test	3.51	.822	H
Listening Test	3.30	.703	M
Writing Test	3.05	.663	M
Overall mean	3.35	.762	M

Note: *VH =Very High, H=High, M=Moderate, L = Low, VL = Very Low

Table 4 shows that the content validity of the AEPET is not in parallel between components. Both Speaking and Reading Tests have high content validity while Listening and Writing Tests have moderate content validity. In other words, the course content is highly represented in Speaking and Reading Tests meanwhile it

is moderately represented in Listening and Writing Tests. The overall mean scores of four English components is 3.35 and this shows that the AEPET on the whole has moderate content validity (M=3.35).

4.3.2 Content validity for AEPET's Preparation

Table 5 Mean Scores for AEPET's Preparation

No	Items	Mean	SD	Degree
1	When constructing the test items, I refer to the test specification.	1.75	.550	L
2	I am given sample AEPET papers before constructing items	3.50	.945	H
3	I refer to English course syllabus when constructing test items	2.30	.923	M
4	The answer keys for the objective questions and marking scheme are prepared before the examination is administered.	3.55	.887	H
5	The questions, answer keys for the objective questions and marking scheme are vetted in a meeting.	1.95	.510	L
6	I attend rater training session (moderation) before assessing the answer scripts assigned to me.	2.25	.786	L
7	Overall mean	2.48	.780	L

Note: *VH =Very High, H=High, M=Moderate, L = Low, VL = Very Low

Table 5 shows that the process of AEPET preparation seems to be rather deficient. The results show that the AEPET seems to be lack of uniformity in referring to relevant documents during the test development. In specific, the lecturers do not refer to the test specification (M = 1.75); the lecturers are uncertain in referring to the English course syllabus (M=2.30); most of the lecturers refer to test sample to decide on the topics to be covered, time allocation, number of test items, difficulty level, and mark allocation (weight age) when constructing AEPET items. Although all the questions, answer keys and marking scheme for the AEPET were prepared before the examination (M= 3.55), they were not put through any quality assurance to remove possible

mistakes or ambiguity in the questions. Furthermore, the crucial vetting session for questions, answer keys and marking scheme is not practiced (M=1.95). Similarly, rater training session for ensuring the consistency in interpreting the marking scheme and judging the performance assessments, especially for AEPET Writing and Speaking components is not carried out (M =2.25). These could have contributed to the moderate content validity of the AEPET, especially the low content validity of AEPET Writing component.

5. Discussion and Conclusion

Overall, the results show that all the AEPET components: Listening, Reading, Writing and Speaking has concurrent,

predictive and content validity. Thus, they can be considered as a valid English language test.

More specifically, for concurrent validity, the AEPET has concurrent validity similar to that of the IELTS, thus students, who achieve higher scores on the AEPET Listening, Reading, Writing and Speaking, have the tendency to obtain similar scores on the same sections of the IELTS. This finding is consistent with several research findings from similar previous studies which focused on investigating the concurrent validity of in-house language tests (Lee, 1995; Liauh, 2011; Nakamura, 2006; Riazi, 2013; Weir, Chan & Nakatsuhara, 2013).

For predictive validity, the AEPET is considered as good predictor for students' academic performance. Therefore, it is suggested that those students, who have high scores on the AEPET, are likely to achieve better CGPA scores. This finding is in line with some researchers (Ajibade 1993; Elder, 1993; Fakeye & Ogunsiji, 2009; Al Hajr, 2014; Maleki & Zangani, 2007; Othman & Nordin, 2013; Sahragard, Baharloo & Soozandehfar, 2011) who found a significant and positive relationship between in house English language tests at university and academic performance, as measured by CGPA and showed that English language tests can predict the students' academic performance.

For content validity, the content validity of the AEPET was investigated in

to two aspects: AEPET's components and preparation. In terms of AEPET components, both Speaking and Reading Tests have high content validity while Listening and Writing Tests have moderate content validity. Overall, the AEPET has moderate content validity. However, the process of AEPET preparation seems to be rather deficient in content validity due to lack of uniformity in referring to relevant documents for constructing the AEPET. Furthermore, the crucial vetting session for questions, answer keys and marking scheme is not practiced. Similarly, rater training session for ensuring the consistency in interpreting the marking scheme and judging the performance assessments, especially for AEPET Writing and Speaking components is not carried out. These could have contributed to the moderate content validity of the AEPET, especially the low content validity of AEPET Writing component. Cunningham, Callahan, and Field (2013) state that the combination of multi-material (specification, textbooks, syllabus) and multi-step internal review helps to design a good test. More importantly, Weir (1993) suggests that in order for test construction to be effective and useful for educators, the test writers should bring all the contents that students have already learned in class in the test. This will provide a good test which produces beneficial washback effect on both teaching and learning. Thus, in order to

increase the content validity of the AEPET, it is necessary to provide the English language lecturers with test specification, course syllabus before test construction. Furthermore, it is a requirement for the lecturers to do the vetting and rater training sessions before the examination is administered. When these suggestions are improved, the strength of correlation for concurrent and predictive validity will increase and make the in-house test be more valid. In other words, it is crucial to re-look at the standard procedure for test development in which the process of test preparation must follow the fundamental test construction guidelines.

REFERENCE

1. Alderson, J. C. (2000). *Technology in testing: the present and the future*. Cambridge: Cambridge University Press.
2. Bachman, L. F., & Palmer, A. (1996). The construct validation of self-ratings of communicative language ability. *Language Testing*, 6 (2), 13-20.
3. Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Philadelphia, U.S: Harcourt Brace Jovanovich College Publishers.
4. Riazi, M. (2013). Concurrent and predictive validity of Pearson Test of English Academic. *Language Testing and Assessment*, 2(2), 1-27.
5. Weir, C. (1993). *Understanding and developing language tests*. New York, U.S: Prentice Hall.

(Toà soạn nhận bài viết ngày 29/8/2017, duyệt đăng ngày 30/9/2017)