

THE PREDICTIVE VALIDITY OF STANDARDISED TEST OF ENGLISH PROFICIENCY IN VIETNAM (VSTEP)

Thieu Ngoc Hung

The People's Police Academy, Vietnam

Email address: ngochungthieu@yahoo.com

<https://doi.org/10.51453/2354-1431/2023/917>

Article info

Received: 20/01/2023

Revised: 17/02/2023

Accepted: 15/03/2023

Keywords:

VSTEP, predictive validity, English language teaching; language testing and assessment

Abstract:

In the field of language testing and assessment, a test's predictive validity can be considered one of the most frequently discussed validity in test validation because it mostly addresses the question of test accuracy as a predictor of academic performance. Hence, this study evaluates the predictive validity of the Vietnamese Standardised Test of English Proficiency (VSTEP) according to the Vietnamese Framework for Foreign Language Competency in order to determine the academic success of students by finding the correlation between VSTEP test scores and their cumulative grade point average (CGPA). The results indicated that there were some existing correlations between VSTEP scores and CGPA; however, these correlations were weak. In precise, the correlation between VSTEP overall score and CGPA had a better predictive value and the VSTEP Writing test showed a bit higher predictive evidence than the other three components: Listening, Reading and Speaking. The study finally covers some recommendations for the use of in-house language tests.



GIA TRỊ DỰ ĐOÁN CỦA BÀI KIỂM TRA ĐÁNH GIÁ NĂNG LỰC TIẾNG ANH CỦA VIỆT NAM (VSTEP)

Thiều Ngọc Hưng

Học viện Cảnh sát nhân dân, Việt Nam

Địa chỉ email: ngochungthieu@yahoo.com

<https://doi.org/10.51453/2354-1431/2023/917>

Thông tin bài viết

Ngày nhận bài: 20/01/2023

Ngày sửa bài: 17/02/2023

Ngày duyệt đăng: 15/03/2023

Từ khóa:

VSTEP, giá trị dự đoán, giảng dạy tiếng Anh; kiểm tra và đánh giá ngôn ngữ

Tóm tắt

Trong lĩnh vực kiểm tra và đánh giá ngôn ngữ, giá trị dự đoán của bài kiểm tra có thể được coi là một trong những độ giá trị được thảo luận phổ biến nhất vì nó chủ yếu giải quyết câu hỏi về độ chính xác của bài kiểm tra như một yếu tố dự đoán kết quả học tập. Do đó, bài nghiên cứu này đánh giá tính giá trị dự đoán của bài kiểm tra đánh giá năng lực tiếng Anh (VSTEP) theo Khung năng lực ngoại ngữ của Việt Nam nhằm xác định sự thành công học tập của học sinh thông qua việc tìm mối tương quan giữa điểm thi VSTEP với điểm trung bình năm học (CGPA). Kết quả chỉ ra rằng điểm VSTEP dường như không phải là yếu tố dự đoán thỏa đáng về thành tích học tập. Kết quả cũng cho thấy chỉ có điểm thi môn đọc VSTEP Writing góp phần giá trị dự đoán thành công trong học tập của sinh viên. Nghiên cứu cũng đề cập đến một số khuyến nghị về việc sử dụng các bài kiểm tra ngôn ngữ.

1. Introduction

In Vietnam, teaching different English activities in the classroom seems not to be successful enough to help students develop their necessary levels of proficiency to pass the exams. Several researchers claim that due to the requirements of English language proficiency, many recent graduates are unable to find employment in overseas organizations because of the fact that the level of communicative competence in English of most Vietnamese employees is still quite low (Van, 2008) [19]. More precisely, less than 10 graduates from a class of sixty students, who are major in English learning, have English language proficiency which meets the requirements for different job positions such as English language teachers, translators, interpreters (Canh & Barnard, 2009, Hu & McGeown, 2020) [4]

[9]. In other words, to get good jobs with high salaries and have high positions in society, students are required to provide the evidence of proficiency in language. As a result, the measurement of proficiency in the English language has become a key aspect (Dimova, 2020) [6].

More significantly, a global testing campaign, particularly standardized English proficiency tests, has been formed as a result of the English language's globalization. The Vietnamese Standardized Test of English Proficiency (VSTEP) which is a new national English exam used in universities and other schools, has progressively grown in importance as a screening tool in Vietnam (MOET, 2015) [10]. By assessing the correlation between the students' VSTEP scores and academic achievement, as determined by their CGPA, this study seeks to determine the predictive

validity of the VSTEP as a predictor of their academic performance. It is hoped that the study will seek out the predictive validity of VSTEP which serves as a useful medium to assess the language benchmark and in other cases to help the students find good opportunities for their future career (MOET, 2017) [11].

Research questions

The study aims to answer two research questions below:

1. To what extent do the VSTEP scores predict students' academic performance, measured by their CGPA?

2. Does any individual VSTEP subskill: Listening, Speaking, Reading and Writing predict students' academic achievement?

2. Literature review

2.1. Predictive validity

Validity has been judged as “indisputably necessary” for any significant test (Weir, 2005). In order to determine if a test is “good” or not, one typically considers its validity. Although the importance of validity in testing and evaluation is generally acknowledged, test developers and academics have varied views on the matter. Bachman and Palmer (1996) [3] gave a compendium of several perspectives on the concept of validity in language testing and assessment. The most fundamental and traditional concept of test validity is the degree to which a test is accurately measuring what it is intended to measure (Bachman & Palmer, 1996; Ghenghesh, 2015; Salgado & Moscoso, 2019; Weir, 2005) [3] [7] [14] [18]. Similarly, Ghenghesh (2015) holds a similar and specific definition of validity as the extent to which test scores reflect students' true level of language knowledge and skills [7].

Nowadays, as language testing has gone up and down through various stages of development in the past century, test validation and types of validity have also undergone changes. Therefore, more special attention must be paid to test validation when one constructs it because the most important consideration in test evaluation is validity which refers to the appropriateness, meaningfulness, and usefulness of a given test as a measure of what it is supposed to measure (Alavi, 2012; Sim et al., 2019) [1] [15]. However, it always comes to the question of what types of evidence and for what types of validity test validators should collect in order to claim the validity of a test. In this regard,

Weir (2005) presents a socio-cognitive framework for test validation. He distinguishes theory-based validity, context validity, scoring validity, and the two external validities, criterion-related validity and consequential validity of a test [18]. The framework provided specific validation portfolios to tests of different macro skills: listening, reading, speaking, and writing. These models are very practical and useful for test developers.

Among these test validity components, predictive validity, which is one type of criterion-related validity, refers to how accurately a test score predicts performance on the criterion measure at some future point in time (Salgado & Moscoso, 2019; Weir, 2005) [14] [18]. When conducting the predictive validity of a test, it always raises a question of how well the English proficiency test scores can accurately predict students' academic success at university (Rethinasamy & Chuah, 2011; Shi et al., 2020) [12] [13]. Therefore, it suggests that a high correlation between language proficiency scores and academic achievement would indicate a high degree of predictive validity (Huang et al., 2021; Wang & Fan, 2020) [8] [17].

Taking these aspects into consideration, this study intends to give more emphasis to investigate the predictive validity of VSTEP in order to find out how significant predictors of students' performance on the test in the context of language test.

2.2. Vietnamese standardised test of English proficiency (VSTEP)

In Vietnam, the Vietnamese Standardized Test of English Proficiency, which is the first standardized English proficiency test, has been organized by the Ministry of Education and Training since 2015. The nearly-three-hour test comprises of four skills: Listening, Speaking, Reading and Writing. The VSTEP is a standardized test designed to measure the English proficiency of undergraduate students in all universities, colleges and schools in Vietnam. The test is based on the Common European Framework of Reference for Languages (CEFR) and be relevant to the context of English use in Vietnam. The test targets from Level 3 (B1) to Level 5 (C1). The students of universities and colleges need to achieve Level 3 which is equivalent to level B1 of the CEFR test. The innovation - setting the VSTEP as a graduation requirement – is expected to have impacts on what and how teachers teach and what and how students learn, thereby promoting students' English skills development (MOET, 2015) [10].

3. Methodology

3.1. Participants

Participants of this study consisted of 550 undergraduate students (N=550) who were studying English at a public university. More specifically, 63.6% (350) of students were females and 36.4% (200) of their academic year of 2019-2022. Before entering the university, these participants had taken the VSTEP as the admission requirement. Table 1 shows the descriptive statistics of the participants.

Table 1. Descriptive statistics of the participants

Participants	Gender			
	Female	%	Male	%
550	350	63.6	200	36.4

3.2. Data Collection

In attempting to establish VSTEP as a predictor of academic performance, 550 students' VSTEP entrance scores, grade point average in the first and second semesters (GPA1, GPA2), and cumulative *grade point average* (CGPA) in the school year 2019-2022 were collected from university database.

3.3. Data analysis

Data analysis was carried out in two stages. The first stage is to provide the results of the student's scores on VSTEP as a whole and the scores on CGPA and this is accompanied by the results of the scores on each component of the VSTEP. The results in the forms of descriptive and also bivariate statistics would be able to provide an overall picture of the performance of the students in all the measures used in the study

For the second stage of the analysis, a stepwise multiple regression was carried out in order to determine which of the VSTEP components proved to be the best predictor of academic achievement as measured by CGPA. Predictor variables such as students' overall VSTEP scores, scores on 4 skills of VSTEP were analyzed against students' GPA1, GPA2 and CGPA which were used as dependent variables.

3.4. Results

3.4.1. Results on students' VSTEP overall band scores

Table 1 shows the distribution of students' levels according to VSTEP overall band scores proposed by Ministry of Training and Education in 2015.

Table1. Distribution of students' VSTEP overall band scores (MOET,2015) [10]

Band score	Frequency (N)	Percent (%)	CEFR-VN level	Grades	General description
0.0-3.5	220	40.1	Non-rating	Non-rating	under-achieving
4.0-5.5	324	58.9	3	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise while travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
6.0-8.0	6	1.90	4	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in their field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with users of the target language quite possible without imposing strain on either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

Band score	Frequency (N)	Percent (%)	CEFR-VN level	Grades	General description
8.5-10	0	0	5	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express themselves fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices

The VSTEP overall band scores distribution for overall band of 550 students is presented graphically in Figure 1 as below:

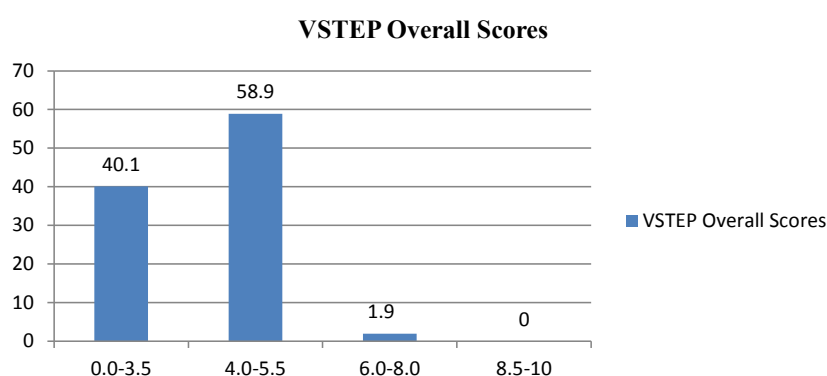


Figure 1. Distribution of students' VSTEP overall scores (N=550)

Figure 1 shows that VSTEP band scores generally range from Band 0-10 which are classified from limited to excellent ability in English proficiency. More specifically, students' overall scores were mainly distributed towards Band 4.0-5.5 (Level 3 – Grade B1), accounting for 58.9%. This signifies that students are able to understand the main points of clear standard input on familiar matters regularly encountered in work, school and leisure time; deal with most situations likely to arise while students are travelling in an area where the language is spoken; produce simple connected text on topics which are familiar or of personal interest; describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans. Then, it is followed by Band 0.0-3.5 with 40.1%, indicating non-rating level and under-achieving English capacity. However, the percentage for Band 6.0-8.0 (Level 4 – Grade B2) was too small, corresponding to 1.9% only and none of them achieved C1 level (Band 8.5-10).

3.4.2. Results on students' VSTEP subtests scores

In order to enrich the status of students' performance on VSTEP, the descriptive statistics results for the VSTEP were calculated. Table 2 shows the means and standard deviations of the VSTEP overall score and subtest scores of 550 students in the school year 2019-2022

Table 2. Descriptive statistics results for the VSTEP scores (N=550)

Score	Min	Max	Mean	SD
VSTEP overall	3.0	6.0	6.184	.7508
VSTEP Listening	2.0	5.0	5.405	1.250
VSTEP Reading	3.0	6.0	5.891	1.131
VSTEP Writing	3.0	6.5	6.297	1.190
VSTEP Speaking	3.0	6.5	6.191	1.244

As can be shown in Table 2, it can be observed that the minimum VSTEP overall score obtained by 550 students was Band 3.0 while the maximum was Band 6.5. On average, the mean score was 6.184 in the VSTEP, showing that the sample as a whole achieved Grade B1 (level 3) and seemed to satisfy the level of English required for academic study at undergraduate at universities.

As for the students' four VSTEP component scores, it can be noted that, the minimum for Reading, Writing and Speaking components got the same Band 3.0 while the lowest Band 2.0 was for listening test. The maximum for both speaking and writing subtests is the highest with Band 6.5, followed closely by reading at Band 6.0. Regarding the averaged mean score, the highest mean score ($M = 6.297$, $SD = 1.190$) was for writing test while the lowest was accounted for listening test with mean score of 5.405. The means for the rest of the components were all above 5.5.

3.4.3. Results on students' academic performance

Table 3 shows the descriptive statistics of students' academic performance

Table 3. Descriptive statistics for GPA1, GPA2 and CGPA (N=550)

Grade point average	Min	Max	Mean	SD
GP1	6.0	8.3	6.923	.8217
GP2	5.8	8.8	7.125	.4316
CGPA	6.0	8.4	6.838	.5525

Table 3 shows that for the first semester, the minimum score was Band 6.0 while the maximum was Band 8.3. On average, the students achieved GPA1 of 6.823. Similarly, the minimum in the second semester was Band 5.8 meanwhile the maximum was Band

8.8. On average, the students scored GPA2 of 7.125. Last but not least, the minimum for the academic performance of the first whole year was Band 6.0 while the maximum was Band 8.4. On the averaged mean score, the students obtained CGPA of 6.838. In Vietnamese grading system, the highest grade is 10.0; therefore the sample with the averaged mean of 6.838 reveals that students as a whole perform standardly well on their academic achievement.

3.4.4. Results of predictive validity of VSTEP

To establish the VSTEP as a predictor of academic performance, the correlation between test scores was calculated. According to George and Mallery (2003), "correlation method is used to examine the relationship of one variable to another" [20, p.12]. Furthermore, they indicated that "correlation is simply a measure of mutual association between two variables" [20, p.14]

As all these references indicated above, Pear Correlation was run to investigate the relationship between the VSTEP overall score and its component scores and CGPA obtained by 550 students. First of all, the correlation coefficient between the VSTEP component scores and CGPA was computed to give a rich descriptive picture of the relationship between the variables (see Table 4)

Table 4. Correlations between the VSTEP components and CGPA (N=550)

	CGPA
VSTEP Listening	.220 ($p=0.137$)
VSTEP Reading	.175 ($p=0.068$)
VSTEP Writing	.365** ($p=0.01$)
VSTEP Speaking	.141 ($p=0.167$)
VSTEP Overall	.372** ($p=.001$)

** Correlation is significant at the 0.01 level (2-tailed)

The remarkable point from Table 4 is that there were insignificant correlations between CGPA and listening ($r=0.220$, $p=0.137$), reading ($r=0.175$, $p=0.068$), and speaking ($r=0.141$, $p=0.167$). However, the correlations between VSTEP writing and CGPA; VSTEP overall score and CGPA have a slightly better predictive value than the others. More specifically, the highest correlation is the correlation between writing and CGPA with the r value of 0.365, then followed closely by the correlation between VSTEP overall and CGPA ($r=0.372$, $p=.001$).

Table 5 shows the correlations between VSTEP scores and GPA1, GPA2 as follows:

Table 5. Correlations between VSTEP scores and GPA₁, GPA₂ (N=550)

	GPA1	GPA2
VSTEP Listening	.042 (<i>p</i> =.671)	.243 (<i>p</i> =.013)
VSTEP Reading	.103 (<i>p</i> =.012)	.101 (<i>p</i> =.045)
VSTEP Writing	.411** (<i>p</i> =.002)	.427** (<i>p</i> =.000)
VSTEP Speaking	.037 (<i>p</i> =.611)	.175 (<i>p</i> =.076)
VSTEP Overall	.302* (<i>p</i> =.001)	.356** (<i>p</i> =.001)

** Correlation is significant at the 0.01 level (2-tailed)

Table 5 generally shows that correlations between each VSTEP component score and GPA1, GPA2 range from the lowest of 0.101 to the highest of 0.427. In specific, high scores on writing test ($r=.411$; $r=.427$, $p=.000$) and overall band scores ($r=.302$; $r=.356$, $p=.000$) tend to associate with high scores on GPA1 and GPA2. However, the association seems to decrease with GPA1 as compared to GPA2. More particularly, none of the significant correlations was found between reading and GPA ($r=.103$; $p=.012$) GPA2 ($r=.101$; $p=.045$) as well as between speaking and GPA1 ($r=.037$, $p=.611$), GPA2 ($r=.175$; $p=.076$)

These results seem to satisfy in light of what Alderson, Clapham and Wall (1995) noticed "... in predictive validity studies, it is common for test developers and researchers to be satisfied when they have achieved a coefficient as low as 0.3" [2, p.182]. Furthermore, Bachman and Palmer (1996) suggest that 0.30 is as high a correlation as can be expected when doing research on the predictive validity. Moreover, a standardized regression coefficient of around 0.3 was common among many of these studies (Sim et al., 2019) [3] [15].

As all these references indicated above, it can be assumed that there is little evidence supporting the predictive validity of the VSTEP in which the correlation between VSTEP overall band score and CGPA; the correlations between Writing test and GPA1, GPA2, CGPA (see Table 5 and 6). All in all, writing test of the VSTEP is only the most significant predictor of students' academic success.

4. Findings and discussion

As all these references indicated above, it can be safely concluded that there are some existing correlations between VSTEP scores and CGPA. However, these correlations are weak. In precise, the correlation between VSTEP overall score and CGPA had a better predictive value than GPA1 and GPA2.

The VSTEP writing component showed a bit higher predictive evidence than the other three components: Listening, Reading and Speaking. This finding suggests that the VSTEP, on the whole, does not predict the students' academic achievement. This is in line with the conclusions drawn from similar previous studies conducted on the predictive validity of various local English language tests (Hu & McGeown, 2020; Salgado & Moscoso, 2019; Teng & Zhang, 2020) [9] [14] [16]. Although these studies varied in the sample sizes, levels of education, and measures of language tests and academic achievement, most of them concluded that the correlation between the language tests and academic achievement was weak to moderate, between 0.2 and 0.4.

From the reasons above, it can be drawn the conclusion that the VSTEP and CGPA have a slightly significant correlation. VSTEP writing component emerged as a satisfactory predictor of students' academic performance as measured by their CGPA. Based on these findings, it suggests that other subtests: Reading, Listening and Speaking should be paid more attention in teaching students who are planning to apply for undergraduate positions at universities or colleges. More prominently, in order to build a more all-inclusive picture of the association between students' English proficiency and their academic accomplishment, educators should take more consideration of other types of test validation such as content validity, face validity, washback which affect on academic performance as well as key influences on the admission criteria.

REFERENCES

- [1] Alavi, T. (2012). The Predictive Validity of Final English Exams as a Measure of Success in Iranian National University Entrance English Exam. *Journal of Language Teaching and Research*, Vol. 3, No. 1, pp. 224-228

- [2] Alderson, J.C., C. Clapham & D. Wall. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press
- [3] *Bachman, L.F. & Palmer, A.S. (1996) Language Testing in Practice Designing and Developing Useful Language Tests*. Oxford University Press, Oxford
- [4] Canh, L.V., & Barnard, R. (2009). Curricular innovation behind closed classroom doors. *Prospect*, 24(2), 20–33.
- [5] Chao, J. Y-G. (2013). Factors affecting college English Foreign Language learners' listening comprehension and listening problems. *NCUE Journal of Humanities*, 8, 71-84.
- [6] Dimova, S. (2020). English language requirements for enrolment in EMI programs in higher education: A European case. *Journal of English for Academic Purposes*, 47, 100896. <https://doi.org/10.1016/j.jeap.2020.100896>
- [7] Ghenghesh, P. (2015). The relationship between English language proficiency and academic performance of university students. *International Journal of Applied Linguistics and English Literature*, 4 (2), 91 - 97.
- [8] Huang, S., Becky, H., Alison, L., Bailey, M., Daniel, A., & Yung-hsiang, S. (2021). An investigation of the validity of a speaking assessment for adolescent English language learners. *Language Testing*, 38(3), 401–428.
- [9] Hu, X., & McGeown, S. (2020). Exploring the relationship between foreign language motivation and achievement among primary school students learning English in China. *System*, 89(1), 21-29. <https://doi.org/10.1016/j.system.2020.102199>.
- [10] Ministry of Education and Training. (2015). Document Application Guidelines Format of test questions to assess English proficiency from level 3 to level 5 according to the 6-level Foreign Language Competency Framework for Vietnam in developing exam questions and marking exams - Approved attached Decision No: 730/QD-BGDDT dated March 11, 2015 of the Minister of Education and Training. Hanoi.
- [11] Ministry of Education and Training. (2017). Decision no. 2080/QD-TTG Approval of the Amendments and Supplements to the Scheme of Teaching and Learning Foreign Languages in the National Education System for the Period of 2017-2025. Hanoi.
- [12] Rethinasamy, S., & Chuah, K. M. (2011). The Malaysian University English Test (MUET) and its use for placement purposes: A Predictive validity study. *Electronic Journal of Foreign Language Teaching*, 8 (2), 234–24.
- [13] Shi, B., Huang, L., & Lu, X. (2020). Effect of prompt type on test-takers' writing performance and writing strategy use in the continuation task. *Language Testing*, 37(3), 361–388. <https://doi.org/10.1177/0265532220911626>
- [14] Salgado, J. F., & Moscoso, S. (2019). The validity of general mental ability for five performance criteria: Hunter and Hunter (1984) revisited. *Frontiers in Psychology*, 10(2), 22-27.
- [15] Sim, F., Thompson, L., Louise, M., Nitish, R., & Wilson, P. (2019). Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review . *PLoS ONE*, 14(2), 22-32. <https://doi.org/10.1371/journal.pone.0211409>
- [16] Teng, L. S., & Zhang, L. J. (2020). Empowering learners in the foreign language classroom: Can self-regulated learning strategies-based writing instruction make a difference?. *Journal of Second Language Writing*, 48(1), 123-134 .
- [17] Wang, L., & Fan, J. (2020). Assessing Business English writing: The development and validation of a proficiency scale. *Assessing Writing*, 46(3), 78-85.
- [18] Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.
- [19] Van, H. V. (2007). The current situation and issues of the teaching of English in Vietnam. Teaching English to speakers of other languages in the internationalization of higher education in Vietnam. *VNU Scientific Journal - Foreign Language*, 3(4), 22-37. 3(4), 1-12.
- [20] George, D., & Mallery, P. (2003). *Statistical package for the social sciences for windows step by step: A simple guide and reference*. 11.0 update (4th ed.). Boston, Massachusetts, U.S: Allyn & Bacon.