

## Language Proficiency, Vocabulary Knowledge, and Individual Factors in Predicting

### Academic Achievement: Insights from the Vietnamese EFL Context

Csaba Z. Szabo<sup>1</sup>, Cao Hong Phat<sup>2</sup>, Barry Lee Reynolds<sup>3</sup>

#### Abstract

The impact of international students' English language proficiency and their vocabulary knowledge on their academic achievement in English-medium degree programmes has been extensively researched, especially in English-immersion contexts. However, given the dearth of research in the Vietnamese EFL context and the restricted predictive power of standardised language proficiency tests compared to vocabulary size, this study contributed to the scholarship of predictive validity by examining the relationships between language proficiency, vocabulary knowledge and academic achievement. A hierarchical linear regression analysis was performed on the PTEA and vocabulary size test (VST) scores of 105 undergraduate students at a teacher-training institution in the South of Vietnam. Findings indicated that language proficiency and vocabulary measures together could moderately predict the academic achievement ( $R^2 = .23$ ) compared to using either language proficiency ( $R^2 = .21$ ) or vocabulary knowledge ( $R^2 = .20$ ) separately. Additionally, partial least squares structural equation modelling (PLS-SEM) indicated a modest predictive power, substantiating the effects of the three indicated individual factors (i.e., autonomous learning activities, language learning attitudes and course perception and evaluation) mediated by language proficiency and vocabulary knowledge on academic

---

<sup>1</sup> University of Edinburgh, <https://orcid.org/0000-0003-1375-0903>

<sup>2</sup> The University of Nottingham, Malaysia Campus; Ho Chi Minh City University of Education, <https://orcid.org/0000-0002-7993-7327>

<sup>3</sup> Faculty of Education, Centre for Cognitive and Brain Sciences, University of Macau, <https://orcid.org/0000-0002-3984-2059>

achievement. These three individual factors were shown to also affect language proficiency ( $R^2 = .31$ ) and lexical knowledge ( $R^2 = .28$ ). This paper offers recommendations for admission and screening procedures of EMI programmes based on multiple measures, such as language proficiency tests, lexical and individual factors.

### **Keywords**

Academic achievement, predictive validity, standardised tests, validation, English as a Foreign Language, English as a Medium of Instruction, partial least structural equation modelling, university admission.

## Introduction

In recent decades, there has been a growing trend of internationalisation in higher education (HE), with English medium of instruction (EMI) programmes gaining popularity worldwide (Knight, 2018; Rose et al., 2021), including Vietnam (Tran & Nguyen, 2018). While EMI programmes can potentially facilitate academic learning, previous research has demonstrated that adequate English language proficiency is prerequisite for students to succeed in academia. (e.g., Daller & Yixin, 2017; Trenkic & Warmington, 2019; Tweedie & Chu, 2019). To determine whether applicants have an adequate linguistic competence, universities mainly rely on results of standardised English language proficiency (ELP) tests such as, the International English Language Testing System (IELTS) Academic, Test of English as a Foreign Language (TOEFL), and Pearson Test of Academic English (PTEA) (Ihlenfeldt & Rios, 2023; Ockey & Gokturk, 2019). The Covid-19 pandemic also gave rise to online, home-based versions of the PTEA and the Duolingo English Test (DET) (Isaacs et al., 2023; Isbell & Kremmel, 2020; Pearson-PTE, n.d.). Despite the PTEA's practicality, previous studies (see Ihlenfeldt & Rios, 2023; Author, year) have called for more validation studies even for the 2011 version.

Standardised ELP test scores predominantly serve as a fundamental criterion for university entry due to their direct relevance to the future academic performance of prospective students in EMI tertiary settings. This is often referred to as predictive validity (Hughes & Hughes, 2020). Mixed findings and weak predictive power regarding the predictive validity of the IELTS and TOEFL have been indicated in the existing literature (e.g., Gagen, 2019; Ihlenfeldt & Rios, 2023). However, little empirical evidence on the predictive validity of the PTEA has been found (also see Ihlenfeldt & Rios, 2023 for details) despite its prominence and wide acceptance.

Compared to IELTS, the PTEA uses the Global Scale of English for a more granular scoring (10-90) that is benchmarked against the Common European Framework for References (CEFR). Arguably, this has the potential to better gauge learners progression and linguistic skills (Pearson, n.d.). Despite this and the introduction of the 2021 off-site version with purportedly improved scoring and shorter duration, which makes it more accessible for international students, no validity evidence is currently available, to our knowledge.

In parallel with predicting academic achievement using ELP tests, vocabulary knowledge has also been claimed as an effective predictor of academic success for students within English-immersion and multilingual contexts (Daller & Yixin, 2017; Milton & Treffers-Daller, 2013; Trenkic & Warmington, 2019; Author, year). When implementing EMI programs in local EFL settings, the importance of vocabulary knowledge becomes even more pronounced. Despite years of English language study, EFL learners have been reported to possess insufficient vocabulary knowledge (Dang et al., 2022; Author, year; Webb & Nation, 2017). This significance of vocabulary is underscored by studies elucidating its correlation with the performance of EMI students across the four language skills, effectively establishing vocabulary as a cornerstone of ELP (e.g., Milton & Treffers-Daller, 2013; Stæhr, 2008). Hence, more research evidence into the predictive power of vocabulary knowledge within EFL contexts is warranted.

Linguistic competence per se cannot account for variations in academic achievement due to its multifaceted nature. Thus, non-language related individual factors also need to be considered; for example, the directionality of the relationships between learning attitudes, learner autonomy and course perceptions have been extensively researched with reference to academic achievement in the existing literature (e.g., Alamer, 2022; Hiver et al., 2021, Sadoughi & Hejazi,

2022). Notably, these studies only focused on English immersion contexts, and therefore much less is known about EFL contexts. To the best of our knowledge, no studies have explored the mediating roles of language proficiency and vocabulary knowledge in the relationships between individual factors and academic achievement.

To contribute the scholarship of predictive validity of the PTEA, the present study examined Vietnamese EMI students' language proficiency (as assessed by overall PTEA scores), vocabulary knowledge (as measured by the Vocabulary Size Test-VST scores), and their individual factors (i.e., autonomous learning activities, language learning attitudes, and course perceptions and evaluations).

## Literature review

### *Language proficiency and academic achievement*

Despite the extensive research into the relationship between ELP, as measured primarily by IELTS and TOEFL, and academic achievement, the findings vary largely across studies. On the one hand, some studies have reported strong correlations ( $r > .60$ ; Plonsky & Oswald, 2014) between standardised ELP tests scores with academic achievement as measured by GPAs (e.g., Bo, et al., 2023; Daller & Phelan, 2013; Harsch et al., 2017; Isaacs et al., 2023; Johnson & Tweedie, 2017). On the other hand, several studies have identified modest to weak associations between ELP tests and academic success (e.g., Bridgeman et al., 2016; Isaacs et al., 2023). Certain studies (e.g., Arcuino, 2013; Arrigoni & Clark, 2015; Isaacs et al., 2023) have indicated no or even negative correlations. These discrepancies in findings indicate the necessity of further investigating the predictive power of ELP tests (e.g., Abunawas, 2014; Gagen, 2019; Ihlenfeld & Rios, 2023).

Meta-studies have highlighted a distinct gap in the number of studies on predictive validity of certain standardised ELP. For example, Ihlenfeld and Rios's (2023) meta-analysis underscored the prevalence of studies centred around the TOEFL, contributing a total of 95 effect sizes. This was followed by IELTS with 26 effect sizes and the PTEA with a mere single effect size. Despite being in operation for over a decade, PTEA has had very limited studies on its predictive validity (i.e., Durrant et al., 2015; Riazi, 2013). Notably, these two studies indicated significant associations between PTEA scores and GPAs compared to the IELTS and TOEL (see Ihlenfeldt & Rios, 2023). Durrant et al.'s (2015) in-house study reported a strong correlation and 68% of the explained variance in GPAs of 81 participants, while Riazi (2013) stated 34% of the explained variance in GPAs of 60 participants. Despite comparable sample sizes, divergent outcomes emerged due to Durrant et al.'s (2015) inclusion of pre-sessional EAP students, while Riazi (2013) encompassed both undergraduate and postgraduate students. Another viable reason is that Durrant et al. (2015) recruited participants with lower proficiency, ranging from A1 to B1 based on the CEFR. It is noteworthy that the PTEA may face challenges in effectively capturing the linguistic gains in Durrant et al.'s (2015) study since the linguistic constructs measured by the PTEA and the modules in the pre-sessional courses might differ. Considering that the PTEA underwent revision in 2021 and no further research has conducted on its predictive validity, further research evidence should be warranted.

### ***Vocabulary knowledge and academic achievement***

Due to the limited predictive power of standardised ELP tests, many researchers have advocated that vocabulary knowledge could contribute significantly to predicting academic performance (e.g., Daller & Xue, 2009; Masrai & Milton, 2017, 2018; Roche & Harrington, 2013; Trenkic & Warmington, 2019). Indeed, an insufficient amount of vocabulary can lead to

learning difficulties, demotivation and even academic failure within university settings (Schmitt, et al., 2017; Trenkic & Warmington, 2019). It is, therefore, required that students should possess a sufficient amount of vocabulary, ranging from 8,000 to 10,000 word families, to successfully perform a range of tasks at the tertiary level (Milton & Treffers-Daller, 2013; Schmitt, et al., 2017; Trenkic & Warmington, 2019). Additionally, administering standardised tests is costly for participants and time-consuming, thereby placing an additional burden on both students and institutions. As such, vocabulary knowledge could contribute significantly to academic performance (e.g., Daller & Xue, 2009; Masrai & Milton, 2017, 2018; Roche & Harrington, 2013; Trenkic & Warmington, 2019).

Having a strong vocabulary is crucial for internalising academic contents in English and, in turn, students with larger vocabularies are likely to perform better in EMI programmes (Nation & Meara, 2019). Past studies have indicated that vocabulary knowledge could better predict academic achievement in various EMI contexts using different vocabulary test formats such as a Yes/No test or an academic vocabulary size test (AVST) compared to standardised ELP tests (see Roche & Harrington, 2013; Masrai & Milton, 2018). On the contrary, several studies found modest or no significant associations between vocabulary measures, such as VST (Nation & Beglar, 2010); VLT (Schmitt et al., 2001) and the Listening Vocabulary Levels Test (LVLT; McLean et al., 2015) and academic achievement (Uchihara & Harada, 2018; Rahman, 2020).

It transpires from these studies that the impact of vocabulary knowledge on academic performance is less clear given the inconsistent findings. Importantly, existing studies employed different measures in different settings (e.g., Japan or Malaysia) leading to varied results and interpretations. In addition, while most studies employed the VST for investigating the

relationship between vocabulary knowledge and students' academic success, it has also faced criticism. The reason for this is that the VST is susceptible to guessing and test taking strategies and for limited number of words tested (Kremmel & Schmitt, 2016; Stoeckel et al., 2021). Still, the VST approximates the written receptive vocabulary size based on the 14,000 most frequent word families of English and thus directly links the VST scores to students' potential reading coverage (Nation & Meara, 2019). Furthermore, mastering a larger vocabulary size from 1,000 to over 8,000 word families becomes a prerequisite as this provides students with the recommended 98% reading coverage for standardised ELP tests (i.e., 2 unknown words in every 100; Laufer & Aviad–Levitzky, 2017). Therefore, exploring how the VST informs EMI tertiary institutions for screening purposes becomes apposite, especially for the Vietnamese EFL context in which students lack high frequency words for comprehending texts in academic texts in English (see Dang et al., 2022; Dang & Webb, 2020).

### *Individual factors and academic achievement*

Academic achievement is a multifaceted construct encompassing a complex amalgamation of cognitive, affective, and behavioural dimensions that reflect an individual's academic progress and proficiency within an educational system (Alamer, 2022; Bo et al., 2023). Due to the multifaceted nature of academic achievement, exploring predictive validity without considering individual factors would not suffice. Thus, considering individual factors alongside linguistic/ELP measures will arguably provide a greater insight into academic achievement.

While most studies mainly focused on the roles of academic disciplines, demographic features and level of study in relation to ELP tests scores or vocabulary knowledge (e.g., Harsch, 2017; Harrington & Roche, 2014, Riazi, 2013), some studies have included a range of individual factors such as learning strategies, educational backgrounds, and previous performances (e.g.,

Avdi, 2011; Durrant, et al., 2015; Ginther & Yan, 2018; Ingram & Bayliss, 2007; Riazi, 2013). These studies generally indicated that individual factors can contribute to students' academic success. In other words, ELP and vocabulary knowledge have been previously researched with respect to academic achievement. However, studies focusing on individual differences, such as language learning attitudes-(LLA), course perceptions and evaluations (CPE) and autonomous learning activities (ALA) have not included them (see Alamer, 2022; Sadoughi & Hejazi, 2022; Hiver, et al., 2021). Furthermore, the roles of ALA, LLA, and CPE in predicting academic achievement with a specific reference to ELP tests and vocabulary knowledge were under-researched (except Durrant et al., 2015; Neumann et al., 2019). The significance of autonomous learning activities and course perceptions of students' academic achievement has been indicated (e.g., Durrant et al., 2015; Neumann et al., 2019) without delving into the nuanced interrelationships between these factors and actual academic success. Consequently, it is critical to investigate the extent to which these individual factors, moderated by linguistic elements, contribute to overall academic achievement.

Although robust methodological approaches, such as structural equation modelling (SEM), can be identified in the prior research, generalising findings to broader populations or making comparisons between studies can be challenging (see Pearson, 2021). The reason for this is that the obtained data in extant research were primarily collected from self-reported questionnaires with different participants, contexts, and varied individual factors.

### **The current study**

The mediating effects of linguistic factors in relation to individual factors on the academic achievement of EMI students in EFL contexts seem underrepresented in the extant literature. Furthermore, there is a notable gap in empirical evidence regarding the predictive

validity of the Pearson Test of English Academic (PTEA), particularly for the revised version. Consequently, the current study aimed to fill this gap and contribute to the empirical validation of the PTEA by investigating the predictive validity of language proficiency, as assessed by PTEA scores on academic performance within an EMI programme at a HE institution within an EFL context. Besides this, the vocabulary size test scores for screening purpose were also of interest as these could potentially inform prospective academic admissions policies and in some cases vocabulary knowledge as measured by the VST can be employed as a component in university entry criteria. Finally, the mediating effects of PTEA and VST scores in relation to ALA, LLA, and CPE on students' GPAs can provide further understanding into the multifaceted nature of academic achievement. Henceforth, we addressed the following research questions:

RQ1: To what extent do undergraduate students' language proficiency (as measured by the PTEA overall scores) and vocabulary knowledge (as measured by the VST scores) predict their academic achievement (as measured by the GPAs)?

RQ2: To what extent are the effects of individual factors (i.e., ALA, LLA, and CPE) on academic achievement (as measured by the GPAs) mediated by language proficiency (as measured the PTEA overall scores) and vocabulary knowledge (as measured the VST scores)?

## **Methodology and research design**

### ***Participants and research setting***

The study was conducted at a public teacher training university in Ho Chi Minh City, Vietnam, using a convenience sampling method that involved 106 EFL Vietnamese undergraduate students (32 male, 74 female,  $M_{age} = 19.10$ ,  $SD = .13$ ). The number of participants recruited for this study exceeds the recommended sample size of 88, which was determined

through a power analysis (Cohen, 2013). Prior to commencing the current study, we conducted a comprehensive power analysis to ascertain the minimum sample size required. Our findings indicated that a sample size of 88 participants would be adequate to detect medium effect sizes, drawing from prior research on the predictive validity of the PTEA (i.e., Durrant et al., 2015; Riazi, 2013). These participants share the same first language and have been studying English for roughly 12 years ( $M = 12.22$ ;  $SD = .25$ ) and enrolled in either Teaching English as a Foreign Language (TEFL) or English Language Studies (ELS) degree programmes.

### *Procedures*

Participants were informed about the study at the beginning of the academic semester. After signing consent forms, participants were guided to register and take the PTEA according to rules and regulations set out by Pearson Education. The PTEA is an internationally standardised and computer-based academic English language proficiency test and adopts automated scoring for all the test components (Pearson-PTE, n.d.). Validation studies have also been conducted by Pearson Education to ensure that its automated scoring technologies produce scores that are comparable to those given by trained human experts, with no significant differences detected for both written and spoken responses (Lu, 2010). All scores produced on the test are measured on the Global Scale of English (GSE), which ranges from 10-90 with 1-point increments (Pearson-PTE, n.d.).

Following the PTEA, the participants completed the VST (Nation & Beglar, 2007) delivered by Qualtrics. The VST is a measure of written receptive knowledge, the form-meaning association and concept knowledge. It comprises 140 multiple-choice items, which are sampled from the 14,000 most frequent word families of English according to the British National Corpus (BNC) frequency list and has been validated using Item-Response-Theory (Beglar, 2010). The

items are dichotomously scored, with 1 point for each correct answer. The total score is then multiplied by 100 to approximate vocabulary size on a scale of 0-14,000.

Subsequently, participants completed the individual factors questionnaire (i.e., ALA, LLA, CPE). The ALA scale involves continuous data while LLA and CPE involved five-point Likert scale questions (see Supplement 1). The values of Cronbach's alpha (.70, .80) and Summability (.24, .52) for LLA and CPE demonstrated acceptable reliability (see Goeman & De Jong, 2018; Larson-Hall, 2015 for details).

Finally, all participants' grade point averages (GPAs) of the English-related modules were retrieved from the Quality and Assurance office of the university after a 15 week-semester. Since this study purported to examine the predictive validity of the PTEA, employing students' GPAs as academic achievement can ensure comparability across studies on predictive validity (Pearson, 2021). Although we acknowledge that GPAs might not be the best measure of academic achievement, it was selected as a composite measure of students' potential attainment of modules and learning outcomes that eventually determine their degree classification. In addition, GPAs are believed to be comparable across different tertiary settings (Pearson, 2021), and thus, were utilised to represent students' academic achievement. According to the Vietnamese HE grading system, GPAs range from 0-10, with 4 being considered a passing score.

### ***Data analyses***

We performed all the statistical analyses using *RStudio 4.2.2*, with the *tidyverse*, *ggplot* and *semnr* packages (Jorgensen et al, 2022; Wickham, et al., 2019). To begin with, the dataset of 106 participants was checked for missing data, outliers, and univariate normality. As the procedure of collecting data was strictly followed, there was no issue of missing data. The data

were subsequently checked for outliers, and values exceeding the interquartile ranges that can affect further statistical analyses were removed (Larson-Hall, 2015). Specifically, the interquartile ranges for PTEA, VST and GPAs were 23.1–82.1, 40–139 and 6.8–9.6, accordingly. One participant was removed as being identified as an outlier for both PTEA and GPAs values (i.e., 10 and 6.5) as its values exceeded the indicated interquartile ranges. Hence, the final sample of this study was 105. Finally, the PTEA overall, VST scores, GPAs and the questionnaire items were subsequently investigated for univariate normality. A guideline of  $\pm 3$  for skewness and  $\pm 10$  for kurtosis was used (Hair et al., 2022) and the obtained results suggested that the data were normally distributed (Supplement 2).

Following the data cleaning, we performed hierarchical linear regression analyses to explore the predictive effects of PTEA overall and the VST scores on GPAs to address RQ1. Correlational patterns were interpreted based on Plonsky and Oswald's (2014) guideline for second language research, whereby correlation coefficients ( $r$ ) of .25, .40 and .60 indicates small, medium and large effect sizes, accordingly. Meanwhile, Plonsky and Ghanbar's (2018) benchmark was employed, with  $R^2 \leq .20$  and  $R^2 \geq .50$  being seen as indicative of small and large respectively in terms of the percentages of explained variance.

We then utilised the partial least squares structural equation modelling (PLS-SEM) to explore the extent to which the individual factors, namely ALA, LLA, and CPE mediated by the PTEA overall and VST scores can predict GPAs. The PLS-SEM best suits the context of this study, which mirrors the modest sample sizes in education and L2 research (Hair & Alamer, 2022). More importantly, the structural model of the current study incorporates formative constructs (i.e., constructs that are defined by their indicators), making the PLS-SEM the preferred method (see Hair, et al., 2022). To assess the certainty of the PLS-SEM for prospective

replicability, a bootstrap routine with 5,000 iterations was employed (Hair et al., 2022). The model was established by assigning ALA, LLA and CPE factors as exogenous variables. Subsequently, the linguistic elements, which comprise the PTEA overall and the VST scores were hypothesised as endogenous constructs and as mediators of the relationships between the individual factors and academic achievement (see Figure 1). Participants' academic achievement was theorised as the outcome variable of the model in this study.

## Results

*To what extent do undergraduate students' language proficiency (as measured by the PTEA overall scores) and vocabulary knowledge (as measured by the VST scores) predict their academic achievement (as measured by the GPAs)?*

### *Preliminary findings*

Table 1 presents the descriptive statistics, distributions, and correlations among all variables, including language proficiency. The PTEA overall ( $M = 52.45$ ,  $SD = 13.72$ ) suggest that most Vietnamese EMI students are around B1 level as referenced by the CEFR, with a moderate degree of variability. Likewise, the VST scores ( $M = 89.92$ ,  $SD = 15.88$ ), showing that the participants' vocabulary sizes disperse around 8,000 word families. Moreover, the PTEA overall showed a strong correlation with the VST scores ( $r = .81$ ,  $p < .01$ ). Similarly, GPAs exhibited positive moderate correlations with both the PTEA overall ( $r = .46$ ,  $p < .01$ ) and the VST scores ( $r = .45$ ,  $p < .01$ ).

**Table 1**

*Descriptive Statistics, Distributions, and Correlations of PTEA overall, VST scores and GPAs*

Variable	<i>S</i>	<i>K</i>	<i>M</i>	<i>SD</i>	Min	Max	1	2
1. PTEA	.23	-.03	52.45	13.72	21.00	87.00		
2. VST	.61	-.10	89.92	15.88	66.00	135.00	.81*	
							[.73, .87]	
3. GPA	.07	-.27	8.23	0.52	7.00	9.60	.46*	.45*
							[.29, .60]	[.29, .59]

*Note.* *S*, *K*, *M*, *SD*, *Min*, and *Max* represent skewness, kurtosis, mean, standard deviation, minimum and maximum values respectively. Values in square brackets indicate the 95% confidence interval for each correlation. \*  $p < .01$

*Primary findings*

To assess the predictive power of language proficiency and vocabulary knowledge, linear hierarchical regressions were performed, with the VST scores being entered in model 1, the PTEA scores in model 2, and both the PTEA and the VST scores in model 3 (see Table 2).

**Table 2**

*Hierarchical linear regression models performance*

Model	Predictor	$R^2$	$R^2$ (adj.)	RMSE	AIC weight	BIC weight	Performance score
1	PTEA	.212	.204	.460	.329	.486	44.15%
2	VST	.207	.199	.462	.230	.340	7.62%
3	PTEA and VST	.231	.216	.455	.441	.173	85.71%

*Note.*  $R^2$ ,  $R^2$ (adj.), RMSE, AIC, and BIC represent R-squared, Adjusted R-squared, Root Mean Squared Error, Akaike Information Criterion weights, and Bayesian Information Criterion weights.

Model 1 included only PTEA overall scores as a predictor of GPAs and 21.2% of the variance in students' GPAs was explained by the PTEA overall scores ( $R^2 = .212$ ). Regarding model 2, the VST scores were added separately as another predictor of GPAs and findings revealed that incorporating the VST scores explained around 20% of total variance in GPAs ( $R^2 = .207$ ) compared to what was explained by PTEA overall scores alone. In the third model, a combination of both the PTEA overall and the VST scores as predictors of GPAs was included, with 23.1% of the variance in GPA being explained ( $R^2 = .231$ ).

The performance of the three models were then checked and compared. The root mean squared error (RMSE) was almost same for all the three models (see Table 2) which represents

the moderate deviation between the predicted and observed values. This suggests that overall accuracy of the three models' predictions is acceptable. The AIC weights were .329, .230, .441 for model 1, 2 and 3, suggesting that model 3 (i.e., having the highest AIC weights) is the best-fitting model among the three. Meanwhile, the BIC weights were .486, .340, and .173 for model 1, 2 and 3, indicating that model 3 (i.e., having the lowest BIC weights) also provides a better goodness of fit and thus is the most appropriate model. The performance score for model 3 ranked the highest (85.71%) while model 2 was the lowest (7.62%) among the three models.

To confirm whether model 3 is significantly better than models 1 and 2, pairwise model comparisons were performed (i.e., model 1 vs. model 3 and model 2 vs. model 3). The results from Table 3 indicate that augmenting the model 1 with VST scores does not significantly improve in GPA prediction ( $p = .114$ ). This suggests that model 1, which exclusively incorporates PTEA, may suffice for explaining GPA, rendering the inclusion of the VST scores unnecessary. Likewise, the comparison between model 2 and model 3 suggests that supplementing the model with the PTEA scores does not better predict GPA ( $p = .074$ ). This implies that model 2 with solely the VST scores, can effectively account for GPA variations, and adding PTEA may not be essential in this context, despite the result being close to significance.

**Table 3**

*Pairwise Model Comparison*

Model	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Model 1	103	22.25				
Model 3	102	21.71	1	.542	2.54	.114
Model 2	103	22.41				
Model 3	102	21.71	1	.694	3.26	.074

*Note.* Res.Df, RSS, Df, Sum of Sq, F, and Pr(>F) are Residual Degrees of Freedom, Residual Sum of Squares, Degrees of Freedom, Sum of Squares, F statistic, and p value of the F statistic.

Overall, model 3 exhibited the highest  $R^2$  value, indicating a better fit to the data and explaining a greater proportion of the variance in the response variable compared to the remaining two models. Model 3 also had the lowest RMSE, suggesting better predictive accuracy. However, findings from pairwise comparisons among the three models revealed that employing the PTEA alone can predict undergraduate students' academic achievement with moderate effect size.

***To what extent are the effects of individual factors (i.e., ALA, LLA, and CPE) on academic achievement (as measured by the GPAs) mediated by language proficiency (as measured the PTEA overall scores) and vocabulary knowledge (as measured the VST scores)?***

*Evaluating the measurement model*

As indicated in the measurement model (i.e., the outer model), three formative constructs involved in the measurement model, including ALA, LLA, CPE. Thus, the measurement model

was evaluated by the convergent validity, which refers to the degree to which an indicator is associated with other indicators under the same construct (Hair et al., 2022). To this effect, a separate redundancy analysis was conducted for each formative construct. The achievement variable (i.e., GPAs) is used as a measure of dependent construct to perform the redundancy analyses for ALA, LLA and CPE (see Table 4). The findings revealed that the path coefficients for ALA, LLA, CPE were .21 ( $R^2 = .05$ ); .38 ( $R^2 = .15$ ); .27 ( $R^2 = .06$ ), suggesting that the three constructs could modestly measure the indicated constructs as referenced against the path coefficient benchmark of .70 ( $R^2 = .50$ ) (Hair et al., 2022). This can be explained by the predominantly *neutral* and *agree* responses obtained for LLA and CPE constructs and very limited amount of time spent on ALA (i.e., 7 hours) via the self-reported questionnaire.

**Table 4**

*Indices of the measurement model*

Constructs	Convergent Validity	Indicator	Outer weight	Outer loading	$R^2$
ALA	$\beta = .21$	ala 1	.18	.25	1.35
		$R^2 = .05$	ala 2	.44	.76
	ala 3		.34	.69	2.01
	ala 4		.34	.44	1.17
	ala 5		60*	-.41	1.15
LLA	$\beta = .38$	lla 1	.40*	.73	1.32
		$R^2 = .15$	lla 2	.63	.82
	lla 3		.26	.47	1.14
	lla 4		-.02	.24	2.15
	lla 5		.33	.29	2.17
	lla 6		-.28	-.03	2.08
	lla 7	.01	.14	2.06	
CPE	$\beta = .27$	cpe 1	-.75	-.14	1.73
		$R^2 = .06$	cpe 2	.64	.58
	cpe 3		.06	.36	3.63
	cpe 4		-.15	.29	4.13
	cpe 5		.36	.44	2.86
	cpe 6		-.48	.09	2.16
	cpe 7	.90	.47	1.97	

*Note.*  $\beta$ ,  $R^2$  and VIF are path coefficients, R-squared and variance inflation factor, respectively.

\*  $p < .1$ , \*\*  $p < .01$

Each indicator of the formative construct provides unique information and is not highly correlated with other items within the construct; thus, the multicollinearity issues in the measurement model were inspected using the variance inflation factor (VIF) of the indicators (Hair, et al., 2022). A VIF value of 5 or higher indicates serious collinearity issues. A VIF below 3 indicates no collinearity, while values between 5 and 3 may be acceptable with theoretical justification (Hair et al., 2022). As shown in Table 4, the results show that the VIF values of the ALA construct ranging from 1.21 to 1.61, LLA from 1.37 to 2.52 and CPE from 1.62 to 4, suggesting that the measurement model does not have any potential issues of collinearity and thus, all indicators are not highly interchangeable.

We then examined the sizes and significances of the indicators of the three constructs. As shown in Table 4, the values of path coefficients ( $w$ ) of all indicators mostly indicate the moderate to strong association between the indicators and their constructs, with lla\_1 ( $p < .01$ ) and lla\_2 ( $p < .10$ ) being significant in bootstrapped weights. However, unexpectedly several indicators show negative correlations with their constructs (i.e., ala\_5; lla\_4; lla\_6, cpe\_1; cpe\_4; cpe\_6), indicating the indicators could not sufficiently represent the hypothesised constructs.

In general, the measurement model exhibits satisfactory results for most indicators, and thus, the structural model was then evaluated.

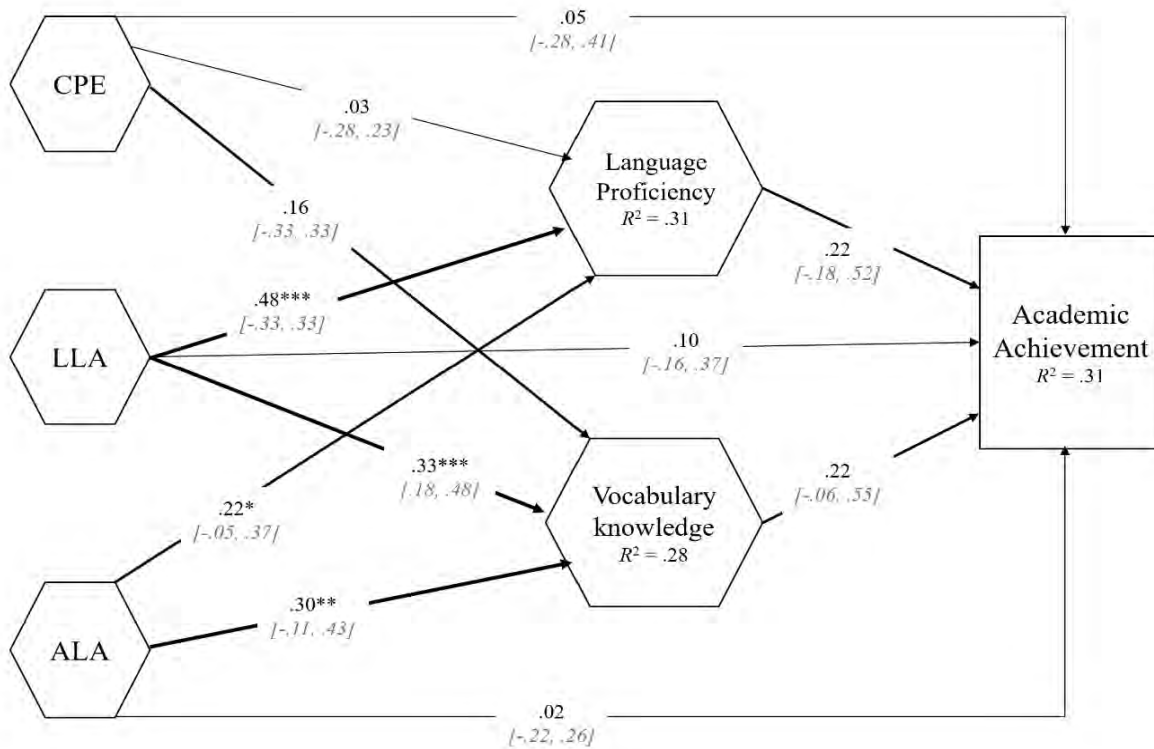
*Evaluating the quality of the structural model*

To start with, the structural model (i.e., inner model) was investigated for collinearity issues and the VIF values for PTEA and VST was below 3.49 and 3.30, respectively. This suggests that the structural model is reliable and valid for examining the relationships among the constructs.

**Figure 1**

*Indices of the structural model*

*Note.* the boldness of the lines indicates the significance levels of the paths; values in brackets are confidence interval (CI) 95%.



Subsequently, standardised path coefficients in the structural model were checked to evaluate the impact of predictor constructs (i.e., ALA, LLA, CPE, language proficiency, vocabulary knowledge) on the dependent variable (i.e., academic achievement) (see Figure). In education and L2 research, path coefficients ( $\beta$ ) are commonly classified as weak, modest, moderate, and strong based on the ranges of 0 to .10, .11 to .30, .30 to .50, and  $> .50$ , respectively. First, the correlations among the three exogenous antecedents of language proficiency were explored. The results indicated positive correlations between ALA, LLA and CPE and language proficiency (i.e., PTEA overall scores) ( $\beta = .22$ , CI 95% [.05, .37];  $\beta = .48$ , CI 95% [.31, .63];  $\beta = .03$ , CI 95% [-.28, .22]), with ALA and LLA being significant. These findings also suggest moderate ( $.11 \leq \beta < .30$ ), modest ( $.30 \leq \beta < .50$ ) and weak effect sizes for ALA, LLA, and CPE, accordingly (Hair et al., 2022). Following this, the relations between the three exogenous antecedents with vocabulary knowledge (i.e., VST overall scores) were then scrutinised, with similar patterns in the findings. In other words, ALA ( $\beta = .30$ , CI 95% [-.11, .44]), LLA ( $\beta = .33$ , CI 95% [.19, .47]) and CPE ( $\beta = .16$ , CI 95% [-.32, .33]) exhibited positive relationships with vocabulary knowledge and moderate effect sizes for ALA, LLA, and CPE were found, with ALA and LLA being significant. When it comes to checking the direct relationships from the three exogenous constructs to the ultimate outcome construct, academic achievement, ALA ( $\beta = .02$ , CI 95% [.22, .25]), LLA ( $\beta = .10$ , CI 95% [-.15, .38]), and CPE ( $\beta = .05$ , CI 95% [-.29, .41]) was also positively associated with the academic achievement (i.e., GPAs) of the participants, exhibiting weak effect sizes for ALA and CPE while moderate effect size for LLA. Finally, the path coefficients of language proficiency and vocabulary knowledge constructs were inspected, with positive correlations being identified. The findings also showed

moderate effect sizes for both language proficiency and vocabulary knowledge constructs (i.e., PTEA:  $\beta = .22$ , CI 95% [-.18, .52], VST:  $\beta = .22$  CI 95% [-.06, .56]).

Once collinearity issues were ruled out and the meaningfulness of the paths was ensured, we assessed the  $R^2$  in the outcome variable to determine the extent to which the predictor constructs explain the variance in the outcomes.  $R^2$  is a commonly used measure in L2 research and is an indicative of in-sample predictive power in PLS-SEM because  $R^2$  is calculated using all the sample data when the algorithm is performed (Hair et al, 2022). In L2 research, it is recommended that  $R^2$  values ranging from 0 to .10, .11 to .30, .30 to .50, and  $> .50$  are generally considered indicative of weak, modest, moderate, and strong explanatory power, respectively (Hair et al., 2022). We focused on the amount of variance explained by the model in predicting the ultimate outcome variable (i.e., academic achievement). It is revealed that the academic achievement has been moderately explained by the predictor variables in the model ( $R^2 = .24$ ). Compared to the results from regression analyses, there is a minor difference between predicting academic achievement by linguistic factors separately and by combining both linguistic and non-linguistic factors ( $R^2 = .24$  vs  $R^2 = .23$ ). Meanwhile, language proficiency ( $R^2 = .31$ ) and vocabulary knowledge ( $R^2 = .28$ ;) variables were moderately and modestly explained by individual factors, respectively. In other words, individual factors could explain 31% and 29% of the variance in the language proficiency and vocabulary knowledge of the participants.

Although the  $R^2$  value is primarily employed to report the predictive power of the model, PLS-SEM practitioners posit that the  $R^2$  is explanatory in nature and is used to provide an in-sample prediction for the model (Hair et al., 2022). To predict novel and prospective observations established based on the samples involved in the model,  $PLS_{predict}$  (i.e., a process for generating out-of-sample prediction) must be used (Hair et al., 2022; Hair & Alamer, 2022). Hair

et al. (2022) suggested comparing the RMSE of the two  $PLS_{predict}$  analyses, namely PLS-SEM and naïve linear regression model (LM) to determine the predictive power for the constructed model. Hence, the RMSE metrics of indicators of language proficiency, vocabulary size and academic achievement (i.e., PTEA overall score, VST overall score and GPAs) in the PLS model (i.e., RMSE values in the PLS in-sample) were then compared with the those in the LM one (LM out-of-sample values).

**Table 5**

*RMSE values of items in the PLS model and the naïve benchmark (out-of-sample) model*

Indicator	PLS model	Naïve benchmark (LM) model
PTEA overall	<b>12.63</b>	8.31
VST scores	<b>14.95</b>	9.88
GPAs	.50	<b>.54</b>

*Note.* Higher prediction error (RMSE) values are in bold.

As illustrated in Table 5, the PLS model had higher RMSE metrics for PTEA overall and VST scores than the LM one (i.e., greater values than the naïve LM benchmark for most indicators). These results suggest more the prediction errors regarding RMSE values. Hence, as referenced against the literature (Hair et al., 2022; Hair & Alamer, 2022), the PLS-SEM model is deemed to have a modest predictive power.

To sum up, the out-sample-prediction value from the PLS-SEM model indicate that the individual factors moderated by the linguistic variables (PTEA overall and VST scores) can

modestly predict Vietnamese undergraduate students' academic achievement. It can be inferred that prospective studies with similar research contexts and designs might generate similar findings. The  $R^2$  metrics indicate that individual factors can moderately predict language proficiency, vocabulary knowledge and academic achievement variables. Additionally, a minor difference was found between the in-sample and out-of-sample prediction power since a moderate variance in Vietnamese undergraduate students' academic achievement was explained by the predictors while the  $PLS_{predict}$  exhibited modest predictive power towards the academic achievement variable.

## Discussion

The present study is the first study, to our knowledge, to examine the predictive validity of the revised version of the PTEA and employed the PTEA overall and the VST scores as moderators of individual factors to predict academic success. This study specifically focused on linguistic elements and individual factors moderated by linguistic factors.

### *Predictive power of language proficiency and vocabulary knowledge*

The VST overall size of 105 Vietnamese EMI students ranges between 6,600 and 13,500 word families ( $M = 90.04$ ;  $SD = 16.76$ ). This suggests that students in this EMI context largely possess the requisite vocabulary. Previous research has recommended a vocabulary size of around 8,000 word families for non-native students to achieve academic success in Anglophone countries (Dang & Webb, 2020; Schmitt et al., 2017; Uchihara & Saito, 2019). While students' PTEA scores revolved around B1-upper and demonstrated a moderate variability, their academic performance, surprisingly, had the least variability, with most students having higher GPAs. Therefore, a score range of B1-upper in the PTEA can be used as a referencing parameter for sustaining an acceptable academic success among undergraduate students as students with that

score range may exhibit consistent academic performance as measured by GPAs whereas lower scores might be used to identify at risk students.

The study revealed a moderate correlation between proficiency test scores and GPAs ( $r = .46$ ). The links among the proficiency test, vocabulary size and academic outcome were corroborated by the PLS-SEM results. While this finding differs from certain previous studies on IELTS and TOEFL that reported weak or negligible relationships (e.g., Arrigoni & Clark, 2015; Bridgeman et al., 2016; Cho & Bridgeman, 2012; Issacs, et al., 2023), it aligns with the consistent pattern of modest yet statistically significant positive associations observed across numerous prior predictive validity studies involving various standardised ELP tests (see Ihlenfeldt & Rios, 2023; Gagen, 2019). This suggests that the updated version of the PTEA can be comparable to other long-standing ELP tests, such as the IELTS and TOEFL.

A strong and significant statistical correlation was identified for the relationship between VST scores and the PTEA ( $r = .81$ ). This is because vocabulary knowledge is fundamental to general language performance (Nation & Meara, 2019). The moderate statistical correlation between from both GPAs and VST scores ( $r = .46$ ) establishes a degree of concurrent validity between these measures. This finding suggests the PTEA and the VST might share similar constructs. Consequently, it is essential to consider additional sources of complementary evidence, for example, vocabulary knowledge alongside standardised ELP tests (Chapelle, 2001) to gain a comprehensive understanding of language proficiency.

With reference to the hierarchical linear regressions, the variance in GPAs explained by the PTEA overall in this study ( $R^2 = .212$ ) is not as promising as those of previous studies on the predictive validity of the PTEA, with Durrant et al., (2015) suggesting 68% explained variance of the GPAs while Razi (2013) 34%. The main difference could also be due to how GPAs are

calculated in these studies, especially it is noticed that GPA in the Vietnamese context lacks variance ( $SD = .52$ ). The observed differences may be attributed to variations in sample size; Durrant et al. (2015) involved 81 participants, while Riazi (2013) involved 64. Moreover, the PTEA employed in this study is the updated version, with shorter timing and different test formats (Pearson-PTE, n.d.). Thus, it is unclear whether the discrepancy is due to the test version or variations in the sample size. Hence, future replication study is imperative. As for the model involving the VST, the result ( $R^2 = .207$ ) is in line with other studies that used lexical tests for predicting academic success, with moderate predictive power being reported (e.g., Harrington & Roche, 2014; Masrai & Milton, 2017, 2018; Rahman, 2020). It can be inferred that the VST in conjunction with the PTEA can be employed for screening purposes in other similar contexts. However, the results of pairwise model comparison suggested that either the VST or the PTEA can adequately capture the academic performance of Vietnamese EMI students. However, considering that the VST is publicly available and lack depth in assessing other language skills, it might be suitable for quick assessments, especially in informal settings or when a rough estimate of vocabulary size suffices. Meanwhile, the PTEA provides a comprehensive evaluation of four language skills that can significantly impact academic outcomes, justifying its role in contexts where in-depth and standardised language assessment is crucial, for example, university admissions.

### ***The mediating effects of language proficiency and vocabulary knowledge***

Drawing on prior research on the impact of individual factors on students' academic outcomes (e.g., Alamer, 2022; Durrant et al., 2015; Neumann et al., 2019; Sadoughi & Hejazi, 2022), a PLS-SEM model was used in the current study. This study is also the first evidence that explored the effects of individual factors (i.e., ALA, LLA, and CPE) on academic achievement

(as measured by the GPAs) mediated by language proficiency (as measured the PTEA overall scores) and vocabulary knowledge (as measured the VST scores). The findings indicated a moderate predictive capacity of individual factors such as autonomy, learning attitudes, and program perceptions on academic achievement. Following Hair et al., (2022), the  $PLS_{predict}$  is mainly dependent on the RMSE and LM in-sample and out-of-sample prediction values. In the context of this study, the inclusion of specific individual variables within the PLS-SEM model yielded a moderately effective predictive model that can be applied to similar research settings. This finding is generally in line with results of most previous studies regarding the predictive validity of standardised ELP tests (e.g., Abunawas, 2014; Bridgeman et al., 2016; Gagen, 2019; Ihlenfeld & Rios, 2023). Moreover, findings demonstrated that incorporating both individual factors and linguistic elements into a predictive model is regarded as the optimal method to research the relationships among several variables rather than solely relying on the correlation and regression analyses. This underscores the notion that academic achievement is a multidimensional construction (Bo et al, 2022; Crowther, 2020; Pearson, 2021), and therefore, a comprehensive range of measures should be considered for informed admission decisions.

The findings from the PLS-SEM model partially echo Durrant et al.'s (2015) conclusion that autonomous learning activities can affect their performance on the PTEA and subsequent academic outcomes despite different statistical methods employed. Hence, understanding the influence of these individual factors takes a multifaceted approach and provides valuable insights into fostering students' academic success. Notably, the effects of these individual factors on the academic outcome were weak in comparison with the impacts of these individual factors (except for CPE) on language proficiency and vocabulary knowledge. Additionally, based on the descriptive statistics for all CPE items in survey questionnaires, the responses were

predominantly positive about their enrolled courses whereas central tendency phenomenon was attributed to the results of the PTEA, the VST and GPAs. In this regard, students' course evaluations are not truly reflected via their linguistic and academic achievements.

### **Limitations and Conclusions**

It is important to acknowledge several limitations of the present study. First, the research's exclusive focus on this demographic raises questions about the generalizability of findings to a broader population. Additionally, as one of the initial investigations utilizing the latest PTEA, the study implies a need for future research to corroborate and extend these preliminary findings. Additionally, the study relied on convenience sampling data, which limits our ability to make causal inferences about the relationships among the variables. Although it is widely accepted that GPAs are seen as the concrete, and thus commonly-used determinant of students' academic success, it may not comprehensively capture all aspects of student performance or achievement. Admittedly, academic success is a multifaceted construct and lacking certain aspects in measuring it via GPAs seems to be inevitable. This is to say that prospective studies can be conducted with other individual factors (i.e., self-efficacy, anxiety, grit, etc.) to better support the predictive validity of the PTEA and support students in achieving academic success.

Despite the indicated limitations, the findings of the present study indicated moderate, but significant relationships between linguistic measures and academic outcomes and demonstrated a modest predictive effect of proficiency test scores on academic success as well. Therefore, HE institutions could provide students with not only language-support programmes but also study skills courses to help students achieve their academic potentials. When it comes to considering individual factors and linguistic components simultaneously, the PLS-SEM model exhibits a

modest predictive power in general. Since the PLS-SEM method is less established in literature of language assessment, it can potentially contribute to the empirical evidence regarding the predictive validity, especially for the revised PTEA.

## References

- Abunawas, M. (2014). *A meta-analytic investigation of the predictive validity of the Test of English as a Foreign Language (TOEFL) scores on GPA* [Unpublished doctoral dissertation]. Texas A&M.  
<https://oaktrust.library.tamu.edu/bitstream/handle/1969.1/154156/ABUNAWAS-DISSERTATION-2014.pdf?sequence=1>
- Alamer, A. (2022). Having a single language interest autonomously predicts L2 achievement: Addressing the predictive validity of L2 grit. *System*, 108, 102850.  
<https://doi.org/10.1016/j.system.2022.102850>
- Arcuino, C. L. T. (2013). *The relationship between the Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS) scores and academic success of international Master's students* [Unpublished doctoral dissertation]. Colorado State University. <https://www.proquest.com/dissertations-theses/relationship-between-test-english-as-foreign/docview/1413309058/se-2>
- Arrigoni, E., & Clark, V. (2015). *Investigating the appropriateness of IELTS cut-off scores for admissions and placement decisions at an English-medium university in Egypt*. (IELTS Research Reports Online Series No. 3). IELTS. <https://www.ielts.org/en-us/for-researchers/research-reports/online-series-2015-3>

- Avdi, E. (2011). IELTS as a predictor of academic achievement in a Master's Program. *English Australia Journal*, 26(2), 42-49.  
<https://search.informit.org/doi/10.3316/informit.361619889813753>
- Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 27(1), 101-118. <https://doi.org/10.1177/0265532209340194>
- Bian, X., Cai, X., & Cai, D. (2021). The Contributions of Listening and Reading Vocabularies to Listening Comprehension of Chinese EFL Students. *International Journal of Listening*, 35(2), 110-122. <https://doi.org/10.1080/10904018.2019.1623678>
- Bo, W. V., Fu, M., & Lim, W. Y. (2023). Revisiting English language proficiency and its impact on the academic performance of domestic university students in Singapore. *Language Testing*, 40(1), 133–152. <https://doi.org/10.1177/02655322211064629>
- Bridgeman, B., Cho, Y., & DiPietro, S. (2016). Predicting grades from an English language assessment: The importance of peeling the onion. *Language Testing*, 33(3), 307-318.  
<https://doi.org/10.1177/0265532215583066>
- Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, 29(3), 421-442. <https://doi.org/10.1177/0265532211430368>
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioural Sciences (2<sup>nd</sup> ed.)*. Routledge.  
<https://doi.org/10.4324/9780203771587>
- Crowther D. (2020). Language investment during university adjustment: The divergent path of two international Chinese freshmen. *Journal of Language, Identity & Education*, 19(4), 275–289. <https://doi.org/10.1080/15348458.2019.1672075>

- Daller, M. H., & Phelan, D. (2013). Predicting international student study success. *Applied Linguistics Review*, 4(1), 173-193. <https://doi.org/10.1515/applirev-2013-0008>
- Daller, M. H., & Xue, H. (2009). Vocabulary Knowledge and Academic Success: a Study of Chinese Students in UK Higher Education. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary Studies in First and Second Language Acquisition: The Interface Between Theory and Application* (pp. 179-193). Palgrave Macmillan UK. [https://doi.org/10.1057/9780230242258\\_11](https://doi.org/10.1057/9780230242258_11)
- Daller, M., & Yixin, W. (2017). Predicting study success of international students. *Applied Linguistics Review*, 8(4), 355-374. <https://doi.org/10.1515/applirev-2016-2013>
- Dang, T., & Webb, S. (2020). Vocabulary and Good Language Teachers. In C. Griffiths & Z. Tajeddin (Eds.), *Lessons from Good Language Teachers* (pp. 203-218). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108774390.019>
- Dang, T. N. Y., Webb, S., & Coxhead, A. (2022). Relationships between lexical coverage, learner knowledge, and teacher perceptions of the usefulness of high-frequency words. *Foreign Language Annals*, 55(4), 1212-1230. <https://doi.org/https://doi.org/10.1111/flan.12663>
- Durrant, P., Walker, C., & Michel, R. (2015). *Research Note: Using PTE Academic to predict achievement and measure proficiency gains in an intensive EAP foundation programme*. Retrieved from [https://assets.ctfassets.net/yqwtwibiobs4/12qmDp6tINQNKurRYZxX9I/63b4a7a686f2ce210a37ce26f7cc3ca1/Using\\_PTE\\_Academic\\_to\\_predict\\_achievement\\_and\\_measure\\_proficiency\\_gains\\_in\\_an\\_intensive\\_EAP\\_foundation\\_programme\\_-\\_Philip\\_Du.pdf](https://assets.ctfassets.net/yqwtwibiobs4/12qmDp6tINQNKurRYZxX9I/63b4a7a686f2ce210a37ce26f7cc3ca1/Using_PTE_Academic_to_predict_achievement_and_measure_proficiency_gains_in_an_intensive_EAP_foundation_programme_-_Philip_Du.pdf)

- Gagen, T. (2019). *The predictive validity of IELTS scores: A meta-analysis* [Unpublished doctoral dissertation]. The University of Western Ontario.  
<https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=8762&context=etd>
- Ginther, A., & Yan, X. (2018). Interpreting the relationships between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Language Testing*, 35(2), 271-295.  
<https://doi.org/10.1177/0265532217704010>
- Goeman, J. J., & De Jong, N. H. (2018). How well does the sum score summarize the test? Summability as a measure of internal consistency. *Educational Measurement: Issues and Practice*, 37(2), 54-63. <https://doi.org/10.1111/emip.12181>
- Hair, J., & Alamer, A. (2022). Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research Methods in Applied Linguistics*, 1(3), 100027.  
<https://doi.org/10.1016/j.rmal.2022.100027>
- Hair, J., Hult, G., Ringle, C., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3<sup>rd</sup> ed.). SAGE. <https://doi.org/10.1007/978-3-030-80519-7>
- Harrington, M., & Roche, T. (2014). Word recognition skill and academic success across disciplines in an ELF university setting. *Papers in Language Testing and Assessment*, 3(2), 76-99.  
[https://arts.unimelb.edu.au/\\_data/assets/pdf\\_file/0009/1770678/Harrington\\_Roche.pdf](https://arts.unimelb.edu.au/_data/assets/pdf_file/0009/1770678/Harrington_Roche.pdf)
- Harsch C., Ushioda E., Ladroue C. (2017). *Investigating the predictive validity of TOEFL iBT® scores and their use in informing policy in a UK university setting* (TOEFL iBT Research



Report No. 30, ETS Research Report No. RR-17-41). ETS.

<https://files.eric.ed.gov/fulltext/EJ1168729.pdf>

Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2021). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, 0(0). <https://doi.org/10.1177/13621688211001289>

Hughes, A., & Hughes, J. (2020). *Testing for language teachers* (3<sup>rd</sup> ed., Cambridge Professional Learning). Cambridge: Cambridge University Press.

<https://doi.org/10.1017/9781009024723>

Ihlenfeldt, S. D., & Rios, J. A. (2023). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, 40(2), 276-299. <https://doi.org/10.1177/02655322221112364>

Ingram, D., & Bayliss, A. (2007). IELTS as a predictor of academic language performance, Part 1. *International English Language Testing System (IELTS) Research Reports*, 7, 1-68.

[https://www.ielts.org/-/media/research-reports/ielts\\_rr\\_volume07\\_report3.ashx](https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report3.ashx)

Isaacs, T., Hu, R., Trenkic, D., & Varga, J. (2023). Examining the predictive validity of the Duolingo English Test: Evidence from a major UK university. *Language Testing*, 40(3), 748–770. <https://doi.org/10.1177/02655322231158550>

Isbell D. R., Kremmel B. (2020). Test review: Current options in at-home language proficiency tests for making high-stakes decisions. *Language Testing*, 37(4), 600–619.

<https://doi.org/10.1177/0265532220943483>

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). semTools: Useful tools for structural equation modeling. R package version 0.5-6. Retrieved from

<https://CRAN.R-project.org/package=semTool>

- Knight, J. (2018). The changing landscape of higher education internationalisation—for better or worse? In *Perspectives on the internationalisation of higher education* (pp. 13-19). Routledge. <https://doi.org/10.4324/9781315160870-3>
- Kremmel, B., & Schmitt, N. (2016). Interpreting Vocabulary Test Scores: What Do Various Item Formats Tell Us About Learners' Ability to Employ Words? *Language Assessment Quarterly*, 13(4), 377-392. <https://doi.org/10.1080/15434303.2016.1237516>
- Larson-Hall, J. (2015). *A Guide to Doing Statistics in Second Language Research Using SPSS and R* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315775661>
- Laufer, B., & Aviad-Levitzky, T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *The Modern Language Journal*, 101(4), 729-741. <https://doi.org/10.1111/modl.12431>
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496. <https://doi.org/https://doi.org/10.1075/ijcl.15.4.02lu>
- Masrai, A., & Milton, J. (2017). Recognition Vocabulary Knowledge and Intelligence as Predictors of Academic Achievement in EFL Context. *TESOL International Journal*, 12(1), 128-142.
- Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, 31, 44-57. <https://doi.org/10.1016/j.jeap.2017.12.006>
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19(6), 741-760. <https://doi.org/10.1177/1362168814567889>

- Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4(1), 151-172.  
<https://doi.org/10.1515/applirev-2013-0007>
- MOET. (2014). *Circular No. 23/2014/TT-BGDĐT on “Regulations on high-quality programs in universities”*. Vietnam: Ministry of Education and Training.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13.
- Nation, P., & Meara, P. (2019). Vocabulary. In N. Schmitt & M. P. H. Rodgers (Eds.), *An Introduction to Applied Linguistics* (3<sup>rd</sup> ed, pp. 48-67). Routledge.  
<https://doi.org/10.4324/9780429424465-3>
- Neumann, H., Padden, N., & McDonough, K. (2019). Beyond English language proficiency scores: understanding the academic performance of international undergraduate students during the first year of study. *Higher Education Research & Development*, 38(2), 324-338. <https://doi.org/10.1080/07294360.2018.1522621>
- Ockey, G. J., & Gokturk, N. (2019). Standardized language proficiency tests in higher education. *Second handbook of English language teaching*, 377-393. [https://doi.org/10.1007/978-3-030-02899-2\\_25](https://doi.org/10.1007/978-3-030-02899-2_25)
- Pearson-PTE. (n.d.). *PTE Academic just got better!* <https://www.pearsonpte.com/articles/pte-academic-just-got-better>
- Pearson, W. S. (2021). The predictive validity of the Academic IELTS test. *ITL - International Journal of Applied Linguistics*, 172(1), 85-120. <https://doi.org/10.1075/itl.19021.pea>

- Plonsky, L., & Ghanbar, H. (2018). Multiple Regression in L2 Research: A Methodological Synthesis and Guide to Interpreting R2 Values. *The Modern Language Journal*, 102(4), 713-731. <https://doi.org/10.1111/modl.12509>
- Plonsky, L., & Oswald, F. L. (2014). How Big Is “Big”? Interpreting Effect Sizes in L2 Research. *Language Learning*, 64(4), 878-912.  
<https://doi.org/https://doi.org/10.1111/lang.12079>
- Rahman, A. (2020). Receptive Vocabulary Size as a Predictor of Undergraduates' Overall Cumulative Grade Point Average With English as Medium of Instruction in Universities. *Asia-Pacific Social Science Review*, 20(3).
- Riazi, M. (2013). Concurrent and predictive validity of Pearson Test of English Academic (PTE Academic). *Papers in Language Testing and Assessment*, 2(2), 1-27.  
[https://arts.unimelb.edu.au/\\_data/assets/pdf\\_file/0010/1771309/2\\_Riazi\\_0.pdf](https://arts.unimelb.edu.au/_data/assets/pdf_file/0010/1771309/2_Riazi_0.pdf)
- Rose, H., Macaro, E., Sahan, K., Aizawa, I., Zhou, S., & Wei, M. (2021). Defining English Medium Instruction: Striving for comparative equivalence. *Language Teaching*, 1-12.  
<https://doi.org/10.1017/S0261444821000483>
- Sadoughi, M., & Hejazi, S. Y. (2023). Teacher support, growth language mindset, and academic engagement: The mediating role of L2 grit. *Studies in Educational Evaluation*, 77,  
<https://doi.org/10.1016/j.stueduc.2023.101251>
- Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, 50(2), 212-226.  
<https://doi.org/10.1017/S0261444815000075>

- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.  
<https://doi.org/10.1177/026553220101800103>
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, 36(2), 139-152.  
<https://doi.org/10.1080/09571730802389975>
- Stoeckel, T., McLean, S., & Nation, P. (2021). LIMITATIONS OF SIZE AND LEVELS TESTS OF WRITTEN RECEPTIVE VOCABULARY KNOWLEDGE. *Studies in Second Language Acquisition*, 43(1), 181-203. <https://doi.org/10.1017/S027226312000025X>
- Tran, L. T., & Nguyen, H. T. (2018). Internationalisation of higher education in Vietnam through English Medium Instruction (EMI): Practices, tensions and implications for local language policies. *Multilingual education yearbook 2018: Internationalization, stakeholders & multilingual education contexts*, 91-106. [https://doi.org/10.1007/978-3-319-77655-2\\_6](https://doi.org/10.1007/978-3-319-77655-2_6)
- Trenkic, D., & Warmington, M. (2019). Language and literacy skills of home and international university students: How different are they, and does it matter? *Bilingualism: Language and Cognition*, 22(2), 349-365. <https://doi.org/10.1017/S136672891700075X>
- Tweedie, M. G., & Chu, M.-W. (2019). Challenging equivalency in measures of English language proficiency for university admission: data from an undergraduate engineering programme. *Studies in Higher Education*, 44(4), 683-695.  
<https://doi.org/10.1080/03075079.2017.1395008>
- Uchihara, T., & Harada, T. (2018). Roles of vocabulary knowledge for success in English-medium instruction: Self-perceptions and academic outcomes of Japanese



undergraduates. *TESOL Quarterly*, 52(3), 564-587.

<https://doi.org/10.1080/03075079.2017.1395008>

Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *The Language Learning Journal*, 47(1), 64-75. <https://doi.org/10.1080/09571736.2016.1191527>

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

<https://doi.org/10.25170/ijelt.v12i1.1458>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>.