IELTS Washback as a High-Stakes Test on Student Learning: A Hierarchical Modelling Study at a Vietnamese University

Nguyen Thanh Minh[1*]

[1] Faculty of Foreign Languages, Van Lang University, Ho Chi Minh City, Vietnam
[*]Corresponding author's email: minh.nt@vlu.edu.vn
[*] https://orcid.org/0000-0001-6441-6732
https://doi.org/10.54855/ijte.25511

Received: 14/08/2024 Revision: 18/12/2024 Accepted: 30/12/2024 Online: 26/01/2025

ABSTRACT

IELTS has been considered a high-stakes test in the Vietnamese context when its results apply to various purposes. Also, many universities in Vietnam have adopted the IELTS test as the requirement for students to graduate, which may negatively affect their learning. This study aimed to investigate these effects by identifying the mechanism between test factors and the washback of the IELTS test on students' psychological and behavioral aspects of learning. With the involvement of 228 participants, quantitative questionnaires were distributed to collect the data. Then, a hierarchical model was established and analyzed by using the PLS-SEM approach. The findings showed students considered the test utility the most important factor, compared to test stakes and difficulty, because, in the research context, the use of the test received more attention from students. Additionally, within the aspects of learning, learning motivation, and restricted learning were significantly influenced by the washback of the test due to the fact that students were motivated to learn for the test. Finally, the study suggested that test factors played significant roles in predicting the washback of the IELTS test on student learning.

Keywords: IELTS; high-stakes test; test factors; student learning; hierarchical modelling

Introduction

Back in the 1980s, when the studies on washback emerged, the focus was mainly on teachers and teaching (Watanabe, 2004). Although students and their learning are directly related to testing and assessment, they are likely to receive less attention from researchers than others (Cheng et al., 2011; Xie & Andrews, 2013). However, washback on students and their learning has recently been paid more attention as a response to address this gap (Sadler, 2016; Xu & Liu, 2018), despite the fact that it remains significant (Cheng et al., 2015). Together, these two directions of research contribute to the descriptions of how to identify washback and reasons for the appearance of the washback

CITATION | Nguyen, T. M. (2025). IELTS Washback as a High-Stakes Test on Student Learning: A Hierarchical Modelling Study at a Vietnamese University. *International Journal of TESOL & Education*, *5*(1), 1-24. DOI: https://doi.org/10.54855/ijte.25511

concept; nevertheless, these findings could not successfully indicate the mechanism of how washback has on both learning and teaching (Cheng et al., 2011; Xie & Andrews, 2013). One main reason for this issue is the employment of the qualitative approach, which could sufficiently identify factors influencing the washback mechanism but not fully explain how strong or weak these factors affect the washback mechanism (Xie, 2015). Also, these studies tend to focus on the impacts of test preparation, which means students are in the test preparation courses. This could be misleading because it is difficult to identify the influence of teaching on learning despite the fact that teaching has been considered the most crucial factor in the result of test preparation courses (Zhan & Wan, 2016). Besides, the effect of test preparation is only an example of washback, which could not reveal the complete picture of washback (Dong, 2020). Hence, the quantitative approach is needed to investigate the washback mechanism more by identifying the statistical relationship among different variables as well as between these variables and washback.

What is more, in terms of washback on specific kinds of tests, the literature suggests that high-skate tests, such as IELTS, are in favor of many researchers, which contributes to the knowledge of how washback works in various settings (Tsagari & Cheng, 2017). IELTS is considered one of the most popular English proficiency tests worldwide (IELTS, 2021) because it offers test users and test takers in non-English-speaking countries simple, easily comprehensible, and clear time-bound evidence of an individual's English proficiency (Pearson, 2019). In the Asian context, including Vietnam, IELTS is metaphoric as a "fever" because it is not only a gateway to study or immigrate overseas but also a gateway to graduation and employment (Nguyen & Nguyen, 2022). Due to the appraisal of society for IELTS, many universities have adopted IELTS as a language standard or requirement for students to graduate (Nguyen, 2023). This raises concerns about how IELTS, as a standard for the English level, affects student learning at the university level (Allen, 2016). However, in terms of the test washback on students' learning, these studies mainly focus on student learning behaviors, such as test preparation, test-taking strategies, or the implementation of the test score etc. Consequently, the student learning psychology, such as motivation and anxiety, is not sufficiently addressed (Nguyen, 2023). Along with the washback on learning psychology, the factors that cause this washback primarily rely on contextual, teacher-related, and learner-related but not the test itself (Nguyen, 2023; Watanabe, 2004).

In Vietnam, implementing the National Foreign Language Project, which aims to reform English teaching, learning, and assessment, considerably impacts society (Bui & Nguyen, 2016). For university students, the English proficiency required after graduating is at least B2 level within the Vietnam Framework of Foreign Language Competency (VFFLC), which equals IELTS 6.0 and above. This creates pressure to learn English for all students and even causes a loss in their learning motivation. Also, for the curriculum developers, the issue would be integrating the majors' specific knowledge and the English program so that students could pass both requirements (Albright, 2018; Tran, 2021). Despite the efforts of equipping university students with higher English proficiency, Vietnamese students still have not met the demands of the work market and society's development (Bui & Nguyen, 2016). Additionally, the complex socioeconomic background and the differences in English input of students make the situation even worse (Ehtsham et al., 2023). Considering all of this, the investigation of the effects of IELTS tests as the graduation requirements on student learning is essential.

Literature review

IELTS as a high-stake test

According to its official website, IELTS refers to the International English Language Testing System, an English proficiency test owned by the British Council (BC) in partnership with IDP Education and Cambridge Assessment English. Indeed, IELTS assesses the English proficiency of test takers through its test components, including reading, speaking, listening, and writing. Also, IELTS is available in two formats: General for those who want to immigrate to English-speaking countries and Academic for those who wish to apply for further education or professional opportunities as the requirements of the host countries (IELTS Homepage). Due to its wide application and its effects on test takers, in most context, IELTS is often viewed as high-stake (Clark et al., 2021).

Regarding the high-stake test, there are several definitions in the literature. Johnson (2008) also indicates that a test is considered high-stake when its test score is used as the gatekeeper for passing or failing students, deciding student graduation, examining teachers' accountability, and schools' image and funding. Share the same perspective, Noori and Mirhosseini (2021) indicate that the outcome of high-stake tests could have crucial effects on test takers. Considering all these definitions, in this study, IELTS as a high-stake test is defined as an English proficiency test which has pivotal impacts on learners, teachers and other stakeholders, and it is considered as a gatekeeper for school admission, graduation, job prospects, or immigrating (Johnson, 2008; Noori & Mirhosseini, 2021).

Effects of testing as impact, washback, and consequence

There are different terms used to describe the impacts of testing, especially high-stake testing, as an educational phenomenon when its results significantly affect the stakeholders related to such testing, including impact, washback, and consequence (Tsagari & Cheng, 2017). Specifically, washback is often used to indicate the effects of testing on learning and teaching in the classroom context (Hughes, 2002). Meanwhile, impact refers to a broader view of the effects of testing as any effects that have on test-takers, policies, or practices within and/or beyond classroom settings such as schools, educational systems, or even society (Wall, 2012). As a result, many researchers and language testers consider washback as one dimension of the impacts of testing (Hamp-Lyons, 1997). What is more, the effects of testing on teaching and learning are usually related to test validity, in which washback needs to be taken into account in measuring test validity as a testing consequence (Messick, 1996). In this study, the effects of testing are examined in relation to student learning; hence, the term washback is employed as the shortened washback of testing.

Natures of washback

Washback is usually identified within five dimensions: specificity, intensity, length, intentionality, and value (Cheng & Watanabe, 2004). *Specificity* refers to the level of specificity of the washback; in other words, washback could be general or specific. Indeed, general washback indicates the effects created by any kind of test, while specific washback presents the impact of a certain type of test or a particular aspect of a test (Alderson & Wall, 1993; Cheng & Watanabe, 2004). *Intensity* describes the power of washback, i.e. the degree of effects in areas of teaching and learning produced by a test (Cheng, 1997). Generally, a high-stake test would affect teaching and learning more than a classroom-based test (Xu & Liu, 2018). *Length* of a washback indicates how long a test's effects

lasts (Watanabe, 2004). *Intentionality* of washback means that washback could be intended or unintended. To be more specific, intended washback refers to effects which test designers expect. In contrast, unintended washback illustrates unexpected effects, such as student anxiety or long-term training on taking tests (Xu & Liu, 2018). Finally, *value* or direction of washback asserts that washback could be positive or negative. For instance, a well-designed and appropriate test could be beneficial to teaching and learning by offering helpful information and creating student motivation. However, a test that is not well-developed or suitable would create negative effects (Xu & Liu, 2018).

Washback models on student learning

Although many conceptual models of washback have been developed, washback studies have indicated that significant variability exists in how teachers, learners, and even other stakeholders have adjusted their behaviors and attitudes towards different testing demands (Ha, 2019). This leads to researchers' struggling to find a common framework to capture this dynamic nature and all the variations of washback (Green, 2013; Liu & Yu, 2021). However, two washback models are repeatedly used by many studies in the field, including the washback trichotomy model by Hughes (1993) and the washback model of student learning by Shih (2007).

The model of Hughes (1993) presents the trichotomy into three main factors, including participants, process and products to construct the basic model of washback. In this model, participants refer to anyone whose work is affected by the perceptions and attitudes towards a test. The process indicates the actions contributing to the learning process, and the product presents what has been learned or acquired (e.g. skills, facts, etc.) and the quality of learning (Hughes, 1993). Also, the mechanism among these is clarified. Testing, first and foremost, affects the participants' perceptions (e.g. teachers and learners) towards their teaching and learning tasks. These perceptions, in turn, impact how these participants conduct their work (process), such as test preparation on similar test items, which will influence the learning outcomes (the products of this work) (Hughes, 1993; Xu & Liu, 2018). In brief, in the model of Hughes (1993), learning behaviors are emphasized.

Regarding the washback model of student learning, Shih (2007) states that the basic model mentioned above could not fully cover washback in social and educational contexts due to the variability of individual learners. Hence, she proposes a model that considers the effects of washback on learning in different psychological aspects, including students' thoughts, experiences, and feelings. She also describes how washback affects students' psychology via extrinsic, intrinsic, and test factors (Shih, 2007).

By examining these two models of washback, this study employs a combination of approaches proposed by Hughes (1993) and Shih (2007), i.e. examining washback on learning in both behavioral and psychological aspects of student learning.

Test factors of IELTS as a high-stake test

Several factors generate washback of a test, including factors related to context, teachers, learners and the test itself (Watanabe, 2004). These proposed factors have been repeatedly examined in different studies. Therefore, in the current study, the test factors are taken into account because there is a limited number of empirical studies on how the test factors influence the washback of the test on student learning (Nguyen, 2023). There are several perspectives on the components of test factors. Watanabe (2004) proposes five components, including test methods, test content, skills

tested, test purpose, test stakes and test status. However, these components partly influence the washback; hence, Shih (2007) and Xie (2015) consider test difficulty to directly affect student learning since students' expectations to pass the test according to its difficulty would enhance their learning. Besides, the implication of the test in reality also impacts students' learning. Specifically, it is often regarded as the extrinsic motivation for student learning (Xie & Andrews, 2013). Bachman (2005) states that the use of tests could be conceptualized as test utility, i.e. the interpretation of the test score and the use of it in learners' situations. Finally, the test had different impacts on learners when it is known as a high-stake or low-stake test. For instance, in the context of not being recognized in the academic transcription, IELTS is viewed as a low-stake test that does not govern students' learning (Stoneman, 2006, as cited in Tsang and Isaacs, 2022). Meanwhile, if the IELTS test plays a role as a graduation gateway, the washback of it is more intensive (Allen, 2016). Based on these reviews, this study viewed test factors as a combination of three aspects: test difficulty, test utility, and test stakes.

Current research on the washback of student learning

The literature has revealed that the effects of washback on student learning are "mixed" between positive and negative (Ha, 2019). The sections below outline the current findings on washback's positive and negative effects.

Positive washback

Some repeatedly cited studies by Xiaoju (1990), Cheng (1998), Hirai and Koizumi (2009), Pan and Newfields (2011), and Allen (2016) show positive washback on student *learning motivation* toward English. Motivation is often considered an essential factor in language learning because it governs and helps maintain students' learning activities as well as influences student engagement and attainment (Schunk, 2012). According to Dörnyei and Ushioda (2021), motivation could be clarified into intrinsic and extrinsic motivation. A test could create intrinsic motivation that students tend to request more extra English materials, participate in activities and read more English materials or journals as well as watch English TV programs (Hirai & Koizumi, 2009; Xiaoju, 1990). Students also spend more time learning English to be better prepared for the test as the results of extrinsic motivation, i.e. to pass the test (Allen, 2016; Cheng, 1998; Pan & Newfields, 2011) and the test result could possibly affect their future prospects (Chu & Yeh, 2017; Nhan, 2013). However, studies by Cheng (1998), Shih (2007) and Pan and Newfields (2012) reveal that tests have minimal effects on learning motivation.

Another positive impact of washback is on student *holistic learning*, i.e., improving skills and abilities or any encouragement to do so due to the test (Alderson & Wall, 1993; Cheng, 1998). indicates that in China's university context, students employ better-coping strategies, such as test-taking, test management and other meta-cognitive strategies to do the test better. Additionally, attempts to do the tests enhance students' abilities and competence significantly (Hung, 2012).

Negative washback

However, washback, at the same time, creates negative effects on student learning. While washback motivates students to learn, it is also responsible for student *learning anxiety* (Shih, 2007). This is a psychological condition occurring when students lack self-confidence in terms of their competence as well as their test results (Al Hadhrami et al., 2024; Tae-Young & Yoon-Kyoung, 2016). This

issue could lead to an adverse impact on their academic performance and to higher levels of stress and depression (Shamsuddin et al., 2013).

Also, washback of testing could lead to *restricted learning*, including rote learning, memorization, past tests, reviewing teachers' notes, etc. (Damankesh & Babaii, 2015; Dong, 2020). Also, students are more likely to employ coping strategies related to the tests rather than cognitive or metacognitive strategies (Xiao, 2014). These lead to some memorization of materials, which could be beneficial to doing the test but not useful for real-life situations or usages (Ren, 2011).

Washback of IELTS on student learning

As mentioned earlier in this section, IELTS is viewed as a kind of high-stake test which greatly impacts several stakeholders, especially students (Johnson, 2008; Madaus & Keillor, 1988; Noori & Mirhosseini, 2021). In the educational context, IELTS plays a role as a tool for measuring students' language ability for enrolment in the academic environment (Pearson, 2019). Due to this nature, in the international context, many researchers have investigated the washback of this test in university settings.

Green (2007) investigated the washback of the IELTS test on students' learning outcomes in terms of their preparation for the test. With a total of 476 participants who took different types of IELTS test preparation courses, the IELTS writing tests and questionnaires were distributed. After using the neural network approach to analyze the data, the author concluded that the course preparation did not affect the scores of the participants. Still, the test difficulty was the determinant of the learners' strategies to prepare for the test.

In another attempt, Stoneman (2006, as cited in Tsang and Isaacs, 2022) compared the impacts of the Graduating Students' Language Proficiency Assessment (GSLPA) and the IELTS on student learning in a Hongkong University. Using a survey and the semi-structured interview as instruments, the study found that these tests significantly affected student motivation to take the preparation course. This could be explained by the awareness of the stakes of these two tests, i.e., an international proficiency exam and a nationwide test. Also, because the IELTS test, in this case, was not included in the transcription, the participants considered its stake to be lower than the other.

Allen (2016) also focused his research on examining the washback of the IELTS test on student test preparation. In his study, 190 participants were involved by completing two IELTS tests, a survey, and an interview for 19 participants. The findings indicated a positive washback in student learning, i.e., improving students' language proficiency with an unequal distribution among the four language skills. This result also varied from those with prior experience in doing the tests and those with high- and low levels of English proficiency. Similar results were also found in the studies of Read and Hayes (2003) and Zhengdong (2009).

In the Vietnamese context, Nguyen (2023) conducted a study on the washback of IELTS on last year English majors' learning at a university. He employed the quantitative research design that followed a combination of frameworks by Hughes (1993) and Shih (2007). Specifically, 282 students participated in his study to respond to a survey on the factors of the IELTS test and how this test affects their learning psychology and behaviors. The collected data were analyzed by using Exploratory Factor Analysis and Confirmatory Factor Analysis with descriptive statistics, Pearson correlation, and multi-linear regression. The findings presented that the IELTS test positively and negatively affected student learning. While IELTS was perceived as important to their study and

future lives, participants considered it hard to reach the required score. Also, the study showed that the mechanism of washback via test factors varied in the research context. Besides these contributions to the knowledge of washback, this study validated a survey for measuring washback of IELTS via psychological and behavioral aspects of student learning, which is adapted in the current study.

From the current literature, the available studies that examined the mechanism of IELTS washback on learning in terms of the test factors remain somewhat inadequate. Also, test factors, which have more influence on student learning psychology and behaviors, still have not received enough attention, especially in the Vietnamese context (Nguyen, 2023). Therefore, this study aimed to examine this mechanism using the hierarchical modelling approach to gain more insights and contribute to the literature on washback studies.

Conceptual framework of the study

From all the reviews above, this study examined the relationship between test factors and the washback of the IELTS test on student learning. Specifically, the test factors are defined into three components: test difficulty, test stakes, and test utility (Allen, 2016; Bachman, 2005; Shih, 2007; Xie, 2015). Also, the IELTS washback on learning are manifested via learning motivation (Cheng, 1998; Hirai & Koizumi, 2009; Pan & Newfields, 2011; Xiaoju, 1990), learning anxiety (Shamsuddin et al., 2013; Shih, 2007), holistic learning (Alderson & Wall, 1993; Cheng, 1998), and restricted learning (Damankesh & Babaii, 2015; Dong, 2020). Figure 1 depicts relationships among these constructs. Additionally, this model has been validated by Nguyen (2023) by using EFA and CFA approaches.

Figure 1

Conceptual framework of the study

Research Questions

Based on the aim stated above and the conceptual framework, three research questions (RQs) have been formulated as follows:

RQ1: How do English majors perceive the factors of the IELTS test in terms of its difficulties, utility, and stakes?

RQ2: How do English majors perceive the impact of the washback of the IELTS test on their learning in terms of their learning motivation, learning anxiety, holistic learning, and restricted learning?

RQ3: Which factors of the IELTS are determinants of the washback of student learning?

Methods

Pedagogical Setting & Participants

The current study was conducted at Vision University (VU, pseudonym), which follows the English requirement of IELTS 6.0 and above for student graduation. The participants in this study were recruited through the convenience sampling method. This kind of non-probability sampling method allows researchers to flexibly contact the participants within their neighborhood or via the Internet (Edgar et al., 2017). In particular, there were 228 participants involved in the study, and the demographic information was presented in the "Findings and Discussion" section.

Design of the Study

This study employed the quantitative approach because the research design provides non-biased and precise measurements, accurately identifying relationship between variables (Mesly, 2015). Also, this design increases the chances of generalizing the results for a larger population (Queirós et al., 2017). The employment of this approach is suitable for the study to investigate the washback mechanism between test factors and student learning.

Research instruments

By conducting a quantitative study, the researcher utilized the questionnaires in collecting data. According to Roopa and Rani (2012), questionnaires allow researchers to collect data from a relatively great number of participants within a short time. Additionally, it offers a more flexible form of distributing via the Internet and in the analysis by using different types of statistical estimation (Creswell, 2014). In this study, the questionnaire was adapted from Nguyen's study (Nguyen, 2023). Table 1 summarizes the constructs and items covered in the instrument.

Table 1

Constructs included in the questionnaire

No.	Constructs		Number of items
1	Demographic information		2
2	Test factors	Test difficulty	9
3		Test stakes	9
4		Test utility	9
5	Washback on student learning	Learning motivation	10
6		Learning anxiety	5
7		Holistic learning	5
8		Restricted learning	5

In short, there were 55 items in the questionnaire, which were divided into three big constructs and 8 sub-constructs: (1) demographic information (gender and year of study), (2) test factors (test difficulty, test stakes, and test utility), and (3) Washback on student learning (learning motivation, learning anxiety, holistic learning, and restricted learning).

Data collection & analysis

The questionnaire was distributed online to the participants via the invitation of the researchers. After the data collection ended, there were 228 responses to the questionnaire, which were collected and coded for data analysis. In the current study, the data analysis procedures followed the PLS-SEM approach and were performed on the SmartPLS 3.0 software. Hair et al. (2013) and Sarstedt et al. (2017) consider PLS-SEM as an effective way of exploring the relationship between exogenous and endogenous variables via the assessment of measurement and structural models with high accuracy. Under this approach, there were three main stages, including (1) establishing the models, (2) assessing measurement models, and (3) assessing the structural model. The detailed descriptions of this analysis procedures were outlined in the "Findings and Discussion" section.

Validity and reliability

Some techniques were employed to ensure the validity and reliability of the study. Firstly, the conceptual framework and the questionnaire were adapted from Nguyen's study (Nguyen, 2023). Moreover, the questionnaire was piloted to 50 respondents before being distributed to the participants. In the pilot process, the reliability of the questionnaire was assessed, using Cronbach Alpha and Composite Reliability estimations. Hair et al. (2013) assert that the reliability of the questionnaire could be assessed via the measurement of Cronbach Alpha (with the value > 0.05) and Composite Reliability (with a value > 0.708). As Table 3.2 shows, all figures for these two estimations met the requirements. As a result, all the items of the questionnaires are reliable.

Table 2

The estimations of Cronbach Alpha and Composite Reliability for the questionnaire pilot

No.	Items	Cronbach Alpha	Composite Reliability
1	Test factors	0.946	0.953
2	Test difficulty	0.911	0.924
3	Test stakes	0.895	0.912
4	Test utility	0.921	0.930
5	Washback on student learning	0.938	0.941
6	Learning motivation	0.894	0.899
7	Learning anxiety	0.848	0.850
8	Holistic learning	0.920	0.922
9	Restricted learning	0.925	0.930

Results/Findings and discussion

Descriptive statistics

In the current study, 228 participants responded to the online survey. The following figures illustrate their demographic information, including gender and their school years.

Figure 2

Gender distribution of the participants

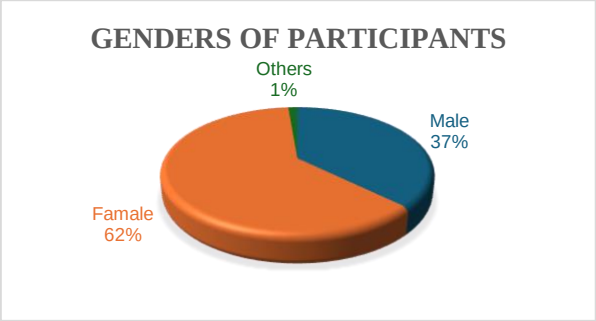
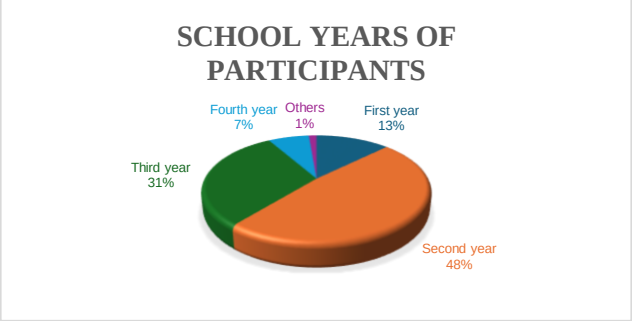


Figure 3

Distribution of school years of participants



Among the participants, the majority of them were female students, with a percentage of 62%, followed by the figures of males (37%) and others (1%). Regarding their years of study, most of them were in their second year (48%) and third year (31%).

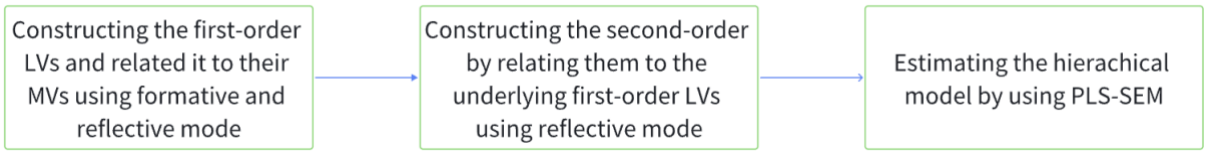
Establishing and assessing the hierarchical model

According to the hypotheses and the conceptual framework proposed after reviewing the current literature, the hierarchical data analysis model was employed to present the relationship between test factors and washback on student learning. Generally, hierarchical models or multidimensional models are considered when researchers aim to investigate constructs with more than one dimension (Crocetta et al., 2021; Heck & Thomas, 2020). In the current study, the latent variable (LV) "test factors" is defined via three manifest variables (MVs), including test difficulty, test stakes, and test utility. Similarly, the latent variable (LV), "washback on student learning", is defined via four manifest variables (MVs), including learning motivation, learning anxiety, holistic learning, and restricted learning. Therefore, two dimensions are involved in the study, forming the hierarchical model's first and second order. This kind of model is beneficial in decreasing the complexity of the model and is more sufficient to utilize the available resources (Crocetta et al., 2021; Heck & Thomas, 2020). Also, the hierarchical model provides high measurement validity (Law et al., 1998).

The PLS-SEM approach is often used to analyze hierarchical models because it offers a tool for examining the relationship and influence of different aspects of a phenomenon (Crocetta et al., 2021). Additionally, the PLS-SEM approach allows the conceptualization of hierarchical models by the repetition of LVs within the model (Guinot et al., 2001). To establish the model using PLS-SEM, this study followed the guidelines from Wetzels et al. (2009) as presented in Figure 4.

Figure 4

Steps to establishing the hierarchical models by using the PLS-SEM approach (adapted from Wetzels et al. (2009))



As Figure 4 suggests, the first step is constructing the first-order LVs and relating them to their MVs using reflective and formative modes. In this study, the first-order models included two sets of MVs, which consisted of seven MVs: test difficulty, test stakes, test utility, learning motivation, learning anxiety, holistic learning, and restricted learning. Then, in the next step, the second-order models of LVs were constructed by relating two LVs – test factors and washback on student learning with their underlying MVs. Specifically, the LV "test factors" was determined by three MVs - test difficulty, test stakes, and test utility – under the formative model. Concerning the LV "washback on student learning", it was manifested by four MVs - learning motivation, learning anxiety, holistic learning, and restricted learning – under the reflective model. After identifying the first and second-order models, the final measurement hierarchical model was proposed by using the PLS-SEM approach, as illustrated in Figure 5. To easily import to the software for analysis, the indicators of all the variables were coded in Table 3.

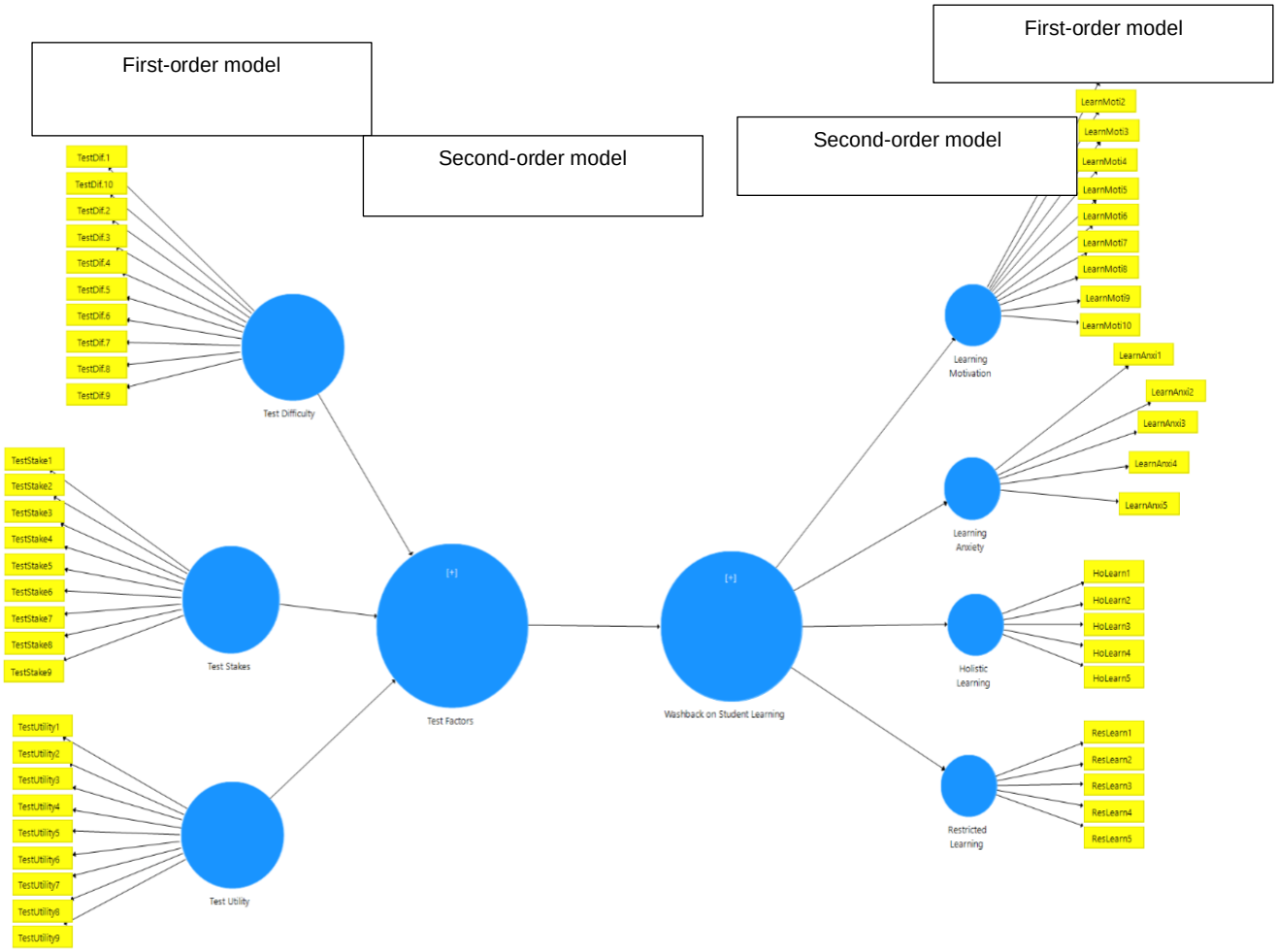
Table 3

Codes used in the data analysis procedures

No.		Constructs	Codes of indicators/ items
1	Gender		Gen
2	Test factors	Test difficulty	TestDif1→ TestDif9
3		Test stakes	TestStake1 → TestStake9
4		Test utility	TestUtility1 → TestUtility9
5	Washback on student learning	Learning motivation	LearnMoti1 → LearnMoti10
6		Learning anxiety	LearnAnxi1 → LearnAnxi5
7		Holistic learning	HoLearn1 → HoLearn5
8		Restricted learning	ResLearn1 → ResLearn5

Figure 5

The hierarchical model of the study



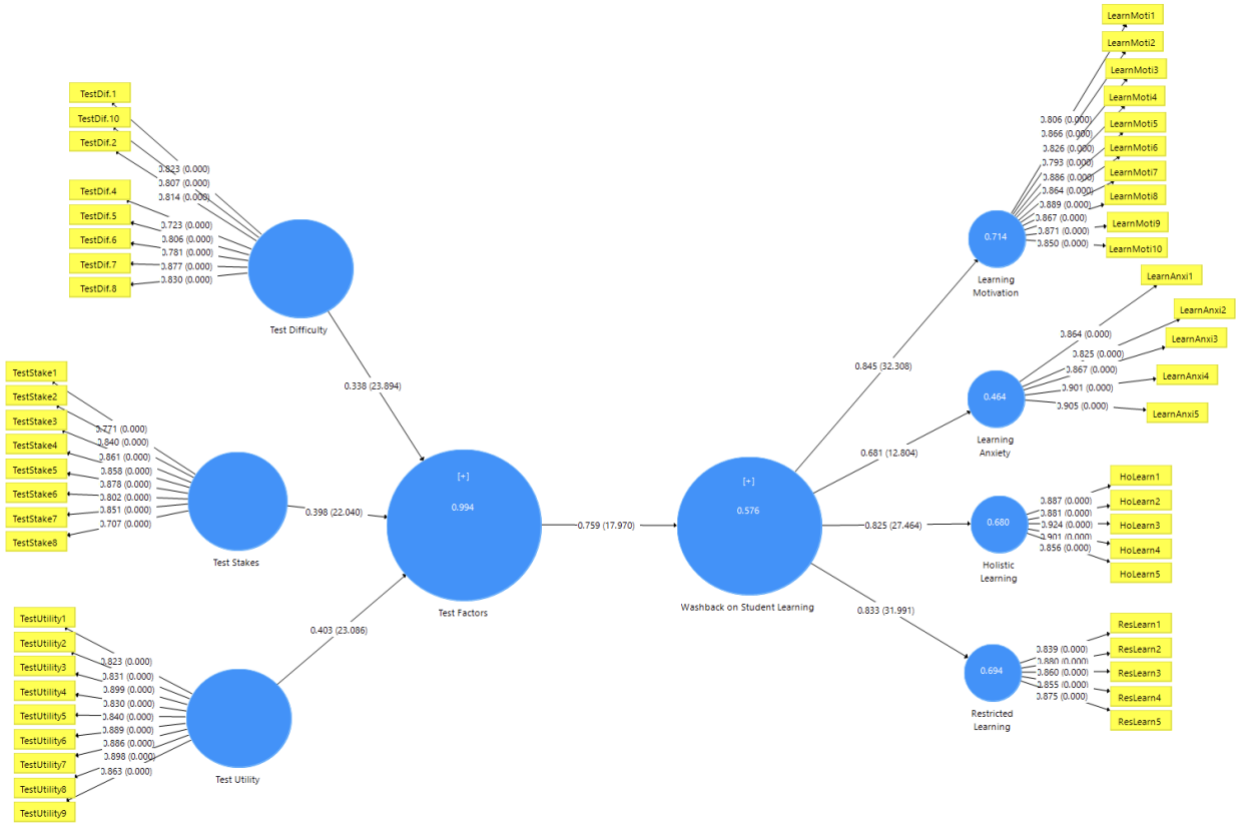
Also, in the final step, the final model was estimated using the PLS-SEM approach in the SmartPLS 3.0. This process was performed within two stages: (1) assessing the measurement model via all the psychometric values of first-order models and (2) assessing the structural model, i.e. the magnitude of the relationships or effects between the variables being considered within the model, including second-order model in Figure 5 (Marcoulides & Saunders, 2006).

Assessing the measurement models

To assess all psychometric values of all nine variables in the model, the outer factor loading, the composite reliability (CR), convergence validity (AVE), and discrimination validity were taken into account (Hair et al., 2013). Firstly, the outer loading factors were estimated to eliminate the indicators that had the value smaller than 0.7 as these did not measure the construct they were supposed to measure within the study's sample. Within this process, the indicators "TestDif3", "TesDif9", and "TestStake9" were removed because in the current estimation, the values of these indicators were below 0.7. As a result, the model with the loading factors is shown in Figure 6.

Figure 6

Hierarchical model with path coefficient



After the removal of all unsatisfied indicators, the remaining values – the composite reliability (CR), convergence validity (AVE), and discrimination validity – were assessed.

Table 4

The psychometric values of the first-order models

Construct	CR	AVE
Test difficulty	0.938	0.654
Test stakes	0.943	0.677
Test utility	0.963	0.744
Learning motivation	0.964	0.726
Learning anxiety	0.941	0.762
Holistic learning	0.950	0.792
Restricted learning	0.935	0.743

In the model, the CR values of the variables in the model must be above 0.78, and AVE must be above 0.5 to ensure that the estimation of the constructs and their relationships are objectives and not impacted by measurement errors (Hair et al., 2019). According to Table 4, all the values of CR and AVE of the variable in the first-order model satisfy this requirement.

Table 5

The HTMT matrix of the first-order models

	Holistic Learning	Learning Anxiety	Learning Motivation	Restricted Learning	Test Difficulty	Test Stakes
Learning Anxiety	0.353					
Learning Motivation	0.755	0.397				
Restricted Learning	0.58	0.613	0.585			
Test Difficulty	0.469	0.55	0.462	0.469		
Test Stakes	0.644	0.473	0.719	0.555	0.607	
Test Utility	0.682	0.418	0.699	0.558	0.581	0.846

In terms of discrimination validity, the HTMT matrix was examined. As Henseler et al. (2015) suggest, all values of the HTMT matrix should be under 0.78 to prove that the examined constructs do not overlap. From Table 5, all the values met the standards.

Assessing the structural model

After ensuring that the first-order models were all valid, Hair et al. (2013) and Sharma and Aggarwal (2019) propose the following statistical estimation should be performed to assess the relationship of the whole model: Collinearity, R^2 explanation of endogenous latent variables, f^2effects size of path coefficients, and Predictive relevance Q^2. Also, at this stage, the nonparametric Bootstrapping was used with 5,000 replications to examine the hypothesis with a significance of 0.05.

Firstly, for collinearity, the VIF was estimated with a standard smaller than the threshold of 3.3 (Roberts & Thatcher, 2009). Via the examination of VIF, the maximum value of the current model was 2.967, which met the standard above. Therefore, in the model, there was no risk of collinearity among variables.

The estimation of R^2 *and* R^2 adjusted are important in determining to what extent the MVs could explain the LVs or in-sample predictive power (Nguyen & Vu, 2020). In the model of the study, the variable Test factors affected the variable Washback on student learning, while the four variables – Holistic learning, Learning anxiety, Learning motivation, and Restricted learning – reflected the variable Washback on student learning. The powers of explanation for these variables are presented in Table 6. According to Henseler et al. (2015), there are three degrees of effect, including low (<0.25), moderate (> 0.25 and < 0.50), and high (>75%).

Table 6

The estimation of R^2 *and* R^2 *adjusted*

	R Square	**R Square Adjusted**	**Comment**
Holistic Learning	0.68	0.679	Moderate effect
Learning Anxiety	0.464	0.461	Moderate effect
Learning Motivation	0.714	0.712	Moderate effect
Restricted Learning	0.694	0.692	Moderate effect
Test Factors	0.994	0.994	High effect
Washback on Student Learning	0.576	0.574	Moderate effect

As Table 6 indicates, most of the variables had moderate effect in in-sample predictive power, except test factors, which could explain Washback on student learning at a high effect degree.

Turning to f^2effects size of path coefficients, the results are shown in Table 7.

Table 7

The estimation of the f^2*effects size of path coefficients*

	P Values
Test Difficulty → Test Factors	0.000
Test Factors → Washback on Student Learning	0.000
Test Stakes → Test Factors	0.000
Test Utility → Test Factors	0.000
Washback on Student Learning → Holistic Learning	0.000
Washback on Student Learning → Learning Anxiety	0.001
Washback on Student Learning → Learning Motivation	0.000
Washback on Student Learning → Restricted Learning	0.000

In Table 7, all the p-values of the f^2effects size of path coefficients were smaller than 0.05. As a result, all the MVs explained the LVs.

Lastly, the Predictive relevance Q^2 was estimated to identify the model's predictive power via the Blindfolding process.

Table 8

The estimation of the Predictive relevance Q^2

	Q^2 (=1-SSE/SSO)
Holistic Learning	0.532
Learning Anxiety	0.345
Learning Motivation	0.509
Restricted Learning	0.509
Washback on Student Learning	0.357

There are two standards for evaluating the Predictive relevance Q^2, including the value of Q^2 and the range of Q^2 value. First of all, if all the Q^2 values of all variables are above 0, then the structural or Hierarchical models in this study reach the global quality (Tenenhaus et al., 2005). In addition, the predictive power of the model is ranked into three levels, including low (the value is between 0 and 0.25), moderate (the value is from 0.25 to 0.5), and high (the value is above 0.5). By examining Table 8, all the figures were nearly over 0.5, which was considered as high predictive power, except the figures for Learning Anxiety and Washback on Student Learning (with moderate predictive power).

Discussion

From the estimation of the current model within the study sample, it is concluded that the proposed model was valid in terms of the measurement of each MV and LV as well as their relationships. The next step was to address the research questions with these findings.

RQ1: How do English majors perceive the factors of the IELTS test in terms of its difficulty, utility, and stakes?

The following table summarizes the result of the model related to the relationship between test difficulty, test stake, test utility and test factors.

Table 9

Total effects of factors of IELTS test

Relationship	Path coefficient	P-value
Test difficulty → Test factors	0.338	0.000
Test stakes → Test factors	0.398	0.000
Test utility → Test factors	0.403	0.000

According to Table 9, the participants in the study considered test difficulty, test stakes and test utility to be highly correlated to test factors (p-value < 0.05). Test utility was also regarded as the most influential factor among these three factors, followed by test skates and test difficulty (the value of path coefficient of 0.403, 0.398, and 0.338, respectively). In the research context, in Vietnamese universities, the IELTS test is highly regarded; however, it is not included in the academic transcriptions. This is in contrast to what Shih (2007) and Xie (2015) found in their research, which found that test difficulty is the most influential factor. Hence, the test stakes in the study did not receive the highest weighting, compared to the utility, which aligns with the finding

of the study conducted by Stoneman (2006, as cited in Tsang and Isaacs, 2022) that the IELTS test was not perceived as important when its stakes were low.

RQ2: How do English majors perceive the impact of the washback of the IELTS test on their learning in terms of their learning motivation, learning anxiety, holistic learning, and restricted learning?

Table 10 presents the relationship between the washback of the IELTS test and different facets of student learning.

Table 10

Total effects of washback on learning of the IELTS test

Relationship	Path coefficient	P-value
Washback on student learning → Learning motivation	0.845	0.000
Washback on student learning → Learning anxiety	0.681	0.000
Washback on student learning → Holistic learning	0.825	0.000
Washback on student learning → Restricted learning	0.833	0.000

As Table 10 suggests, the washback of the IELTS test had a positive relationship with students' learning aspects, including learning motivation, learning anxiety, holistic learning, and restricted learning (p-value <0.05 and values of coefficient> 0). These results clearly illustrated the fact that washback of the test statistically significantly influences learning behaviors and psychology of student learning. This is confirmed in the study of Nguyen (2023) when both psychological and behavioral aspects of student learning were considered. Regarding each aspect in isolation, the effects of the test on students' learning motivation were found by the research of Xiaoju (1990), Cheng (1998), Hirai and Koizumi (2009), and Pan and Newfields (2011), and Allen (2016), while these impacts on students' holistic learning were also identified by Alderson and Wall (1993), Cheng (1998) and Xiao (2014). Specifically, the washback of the test increased learners' motivation to learn English, and in this study, the washback of the IELTS test highly affected learning motivation (with a coefficient of 0.845). This finding was on the contrary with the minimal influence of the test washback and motivation the research by Cheng (1998), Shih (2007), and Pan and Newfields (2012).

Holistic learning was viewed as the positive aspect of learning that the test had washback on (Alderson & Wall, 1993; Cheng, 1998). In this study, the participants perceived this aspect in a lower manner than that of restricted learning, which shared the same results as the study of Xiao (2014), Damankesh and Babaii (2015), and Dong (2020) when learners focused more on the strategies to cope with the test. However, the findings still indicated that learners saw that their English skills improved during the test preparation. Another negative effect of washback on learning is learning anxiety (Shih, 2007). The current study revealed that the test did not correlate with anxiety as highly as the other learning aspects. This finding is in contrast with the studies of Shamsuddin et al. (2013) when the anxiety of students was found to be significant.

RQ3: Which factors of the IELTS are determinants of the washback of student learning?

Table 11 shows the effects of test factors on the test washback of students' learning.

Table 11

Total effects of test factors on the test washback of students' learning.

Independent variable	Dependent variable	Path coefficient	P-value
Test difficulty	Washback on student learning	0.255	0.000
Test stake	Washback on student learning	0.303	0.000
Test utility	Washback on student learning	0.305	0.000

According to Table 11, the components of test factors, including test difficulty, test stakes, and test utility, had statistically significant effects on students learning, known as test washback (path coefficient> 0, and p-value < 0.05). Within the study samples, the test utility had the most influence on the test washback on learning, followed by the test stakes and test difficulty. For the participants of this study, the test usages play the most significant role in their learning English. This reflects the assumption of Pearson (2019) and the social context of Vietnam (ThanhNienNews, 2022) that the IELTS test provides a tool for measuring English proficiency for various purposes, from academic to employment. Also, test stakes impacted student learning in the second position with just a small loading factor of 0.02. This aligned with the conclusion of Stoneman (2006, as cited in Tsang and Isaacs, 2022) and Allen (2016) that the stakes of a test are considered as high when its usages are significant and directly affect the test taker. However, the test difficulty did not affect the students' learning as much as the other. This could be explained by the common knowledge and awareness of the test taker when the IELTS test becomes dominant within international English proficiency tests (IELTS, 2021). Compared to the study of Nguyen (2023), this study found that all test factors had considerable influences on learning aspects, not just some aspects such as learning motivation or restricted learning.

In addition to these findings, the current study contributed to the validation of the washback model based on test factors. The estimation of Predictive relevance Q^2 in the above section indicates that the test factors strongly predicted the impact of the IELTS test on students' holistic learning, learning motivation and restricted learning (Q^2 value > 0.5). Hence, test factors should be considered as an important determinant of the test washback, as suggested by Shih (2007), Xie (2015) and Nguyen (2023).

Conclusion

The current study aimed at investigating the mechanism of how test factors of the IELTS test (including test difficulty, test stakes, and test utility) affect the washback on student learning in terms of learning motivation, learning anxiety, holistic learning, and restricted learning. With the involvement of 228 students at Vision University (VU) and the employment of a survey and PLS-SEM approach in analyzing the hierarchical model, the study indicated that test factors played a significant role in the mechanism of washback on student learning. Specifically, among three test factors, the learners in the research context considered test utility the most important one, followed by test stakes and difficulty. Regarding student learning, learning motivation and restricted learning were recorded as the most affected aspects by the test washback. Regarding the relationship between test factors and test washback on student learning, test utility and test stakes were the determinants compared to test difficulty. The hierarchical model proposed in the study yielded the predictive

power of the models with the examination of test factors, and both psychological and behavioral aspects of learning was relatively high. From the results of the current study, it is apparent that in the context of Vietnam and some other familiar contexts in Asia, the IELTS test is used for various purposes, which indicates the importance of its utility. However, this test is not included in students' transcription; therefore, the stakes of it is low. Additionally, due to its popularity among students, the test format is somewhat familiar to them. As a result, among the test factors, test difficulty is not viewed as the most influential. Finally, the study confirms the possible impacts of high-stake tests on both psychological and behavioral aspects of student learning.

Although the study contributes to the understanding of the washback mechanism of test factors on both learners' psychological and behavioral aspects of learning, there still remained some limitations. First of all, the sample size and sampling methods employed in the study could reduce the generalization of the findings. However, in the study, utilizing the bootstrapping technique to analyze hierarchical models could minimize that risk. Therefore, it is recommended that other researchers could use probability sampling methods with a larger sample size for their studies. Additionally, due to the time constraint, the research design relied on quantitative research design. Nevertheless, as stated in the introduction section, most studies on washback of testing on learning employ qualitative research design. Hence, this study attempted to bridge this gap.

References

Al Hadhrami, I., Al Sawafi, A., Abraham, J., Al Aamri, K., & ALzeidi, A. (2024). Test anxiety: perceptions of Omani nursing students–a descriptive qualitative study. *Nurse Education in Practice*, *76*(1), 10328-10335. https://doi.org/10.1016/j.nepr.2024.103928

Albright, J. (2018). *English tertiary education in Vietnam*. Routledge.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied linguistics*, *14*(2), 115-129. https://doi.org/10.1093/applin/14.2.115

Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, *6*(1), 7. https://doi.org/10.1186/s40468-016-0030-z

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly: An International Journal*, *2*(1), 1-34. https://doi.org/10.1207/s15434311laq0201_1

Bui, T. T. N., & Nguyen, H. T. M. (2016). Standardizing English for Educational and Socio-economic Betterment- A Critical Analysis of English Language Policy Reforms in Vietnam. In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 363-388). Springer International Publishing. https://doi.org/10.1007/978-3-319-22464-0_17

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and education*, *11*(1), 38-54. https://doi.org/10.1080/09500789708666717

Cheng, L. (1998). Impact of a Public English Examination Change on Students' Perceptions and Attitudes toward Their English Learning. *Studies in Educational Evaluation*, *24*(3), 279-301. https://doi.org/10.1016/S0191-491X(98)00018-2

Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, *28*(2), 221-249. https://doi.org/10.1177/0265532210384253

Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, *48*(4), 436-470. https://doi.org/10.1017/S0261444815000233

Cheng, L., & Watanabe, Y. (2004). *Washback in language testing: Research contexts and methods*. Routledge.

Chu, H.-y., & Yeh, H.-n. (2017). English Benchmark Policy for Graduation in Taiwan's Higher Education: Investigation and Reflection. *Journal of Language Teaching and Research*, *8*(6), 1063-1072. https://doi.org/10.17507/jltr.0806.06

Clark, T., Spiby, R., & Tasviri, R. (2021). Crisis, collaboration, recovery: IELTS and COVID-19. *Language Assessment Quarterly*, *18*(1), 17-25. https://doi.org/10.1080/15434303.2020.1866575

Creswell, J. W. (2014). *Educational research: Planning, conducting, and evaluating quantitative*. Pearson Education Limited.

Crocetta, C., Antonucci, L., Cataldo, R., Galasso, R., Grassia, M. G., Lauro, C. N., & Marino, M. (2021). Higher-order PLS-PM approach for different types of constructs. *Social Indicators Research*, *154*, 725-754. https://doi.org/10.1007/s11205-020-02563-w

Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in Educational Evaluation*, *45*, 62-69. https://doi.org/10.1016/j.stueduc.2015.03.009

Dong, M. (2020). Structural relationship between learners' perceptions of a test, learning practices, and learning outcomes: A study on the washback mechanism of a high-stakes test. *Studies in Educational Evaluation*, *64*, 100824. https://doi.org/10.1016/j.stueduc.2019.100824

Dörnyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation*. Routledge.

Edgar, T., Manz, D., & Manz, D. (2017). Exploratory study. *Research methods for cyber security*, *29*, 95-130.

Ehtsham, M., Zeb, A., Alam, A., & Ahmad, S. (2023). Examining the Impact of Socioeconomic Factors on English Language Proficiency: A Cross-Cultural Analysis. *Al-Mahdi Research Journal (MRJ)*, *5*(2), 126-140.

Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre‐sessional language courses. *Assessment in Education*, *14*(1), 75-97. https://doi.org/10.1080/09695940701272880

Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, *13*(2), 39-51. https://doi.org/10.6018/ijes.13.2.185891

Guinot, C., Latreille, J., & Tenenhaus, M. (2001). PLS path modelling and multiple table analysis. Application to the cosmetic habits of women in Ile-de-France. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 247-259. https://doi.org/10.1016/S0169-7439(01)00163-0

Ha, N. T. T. (2019). A literature review of washback effects of assessment on language learning. *Ho Chi Minh City Open University Journal of Science - Social Sciences*, *9*(2), 3-16. https://doi.org/10.46223/HCMCOUJS.soci.en.9.2.257.2019

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long range planning*, *46*(1-2), 1-12.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European business review*, *31*(1), 2-24. https://doi.org/10.1108/EBR-11-2018-0203

Hamp-Lyons, L. (1997). Washback, impact and validity: ethical concerns. *Language Testing*, *14*(3), 295-303. https://doi.org/10.1177/026553229701400306

Heck, R., & Thomas, S. L. (2020). *An introduction to multilevel modeling techniques: MLM and SEM approaches*. Routledge.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1108/F-11-2014-0081

Hirai, A., & Koizumi, R. (2009). Development of a practical speaking test with a positive impact on learning using a story retelling technique. *Language Assessment Quarterly*, *6*(2), 151-167. https://doi.org/10.1080/15434300902801925

Hughes, A. (1993). Backwash and TOEFL 2000. *Unpublished manuscript, University of Reading*.

Hughes, A. (2002). *Testing for Language Teachers* (2 ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511732980

Hung, S.-T. A. (2012). A washback study on e-portfolio assessment in an English as a Foreign Language teacher preparation program. *Computer Assisted Language Learning*, *25*(1), 21-36. https://doi.org/10.1080/09588221.2010.551756

IELTS. (2021). *New milestone for world-leading English test: IELTS trusted by more than 11,000 organisations across the globe*. Retrieved 10 May 2024 from https://www.ielts.org/news/2021/ielts-trusted-by-more-than-11000-organisations-across-the-globe

Johnson, D. D. (2008). *Stop high-stakes testing: An appeal to America's conscience*. Rowman & Littlefield.

Law, K. S., Wong, C.-S., & Mobley, W. M. (1998). Toward a taxonomy of multidimensional constructs. *Academy of management review*, *23*(4), 741-755. https://doi.org/10.5465/amr.1998.1255636

Liu, X., & Yu, J. (2021). Relationships between learning motivations and practices as influenced by a high-stakes language test: The mechanism of washback on learning. *Studies in Educational Evaluation*, *68*, 100967. https://doi.org/https://doi.org/10.1016/j.stueduc.2020.100967

Madaus, G. F., & Keillor, G. (1988). The influence of testing on the curriculum. *Teachers college*

record, *89*(5), 83-121. https://doi.org/10.1177/016146818808900505

Marcoulides, G. A., & Saunders, C. (2006). Editor's comments: PLS: a silver bullet? *MIS quarterly*, iii-ix. https://doi.org/10.2307/25148727

Mesly, O. (2015). Simulation and Quantitative Techniques. In O. Mesly (Ed.), *Creating Models in Psychological Research* (pp. 53-61). Springer International Publishing. https://doi.org/10.1007/978-3-319-15753-5_7

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241-256. https://doi.org/10.1177/026553229601300302

Nguyen, H., & Vu, T. (2020). *Phân tích dữ liệu áp dụng mô hình PLS-SEM [Data analysis using PLS-SEM model]*. NXB Kinh Tế Thành Phố Hồ Chí Minh.

Nguyen, H. N., & Nguyen, D. K. (2022). Vietnamese Learners' Performance in The IELTS Writing Task 2: Problems, Causes, and Suggestions. *International Journal of TESOL & Education*, *2*(1), 170-189. https://doi.org/10.54855/ijte.222111

Nguyen, H. T. M. (2023). The Washback of the International English Language Testing System (IELTS) as an English Language Proficiency Exit Test on the Learning of Final-Year English Majors. *TESL-EJ*, *27*(2), n2. https://doi.org/10.55593/ej.26106a8

Nhan, T. (2013). The TOEIC® test as an exit requirement in universities and colleges in Danang City, Vietnam: Challenges and impacts. *International Journal of Innovative Interdisciplinary Research*, *2*(1), 33-50.

Noori, M., & Mirhosseini, S.-A. (2021). Testing language, but what?: Examining the carrier content of ielts preparation materials from a critical perspective. *Language Assessment Quarterly*, *18*(4), 382-397. https://doi.org/10.1080/15434303.2021.1883618

Pan, Y., & Newfields, T. (2011). Teacher and student washback on test preparation evidenced from Taiwan's English certification exit requirements. *International Journal of Pedagogies and Learning*, *6*(3), 260-272. https://doi.org/10.5172/ijpl.2011.6.3.260

Pan, Y.-C., & Newfields, T. (2012). Tertiary EFL Proficiency Graduation Requirements in Taiwan: A Study of Washback on Learning. *Electronic Journal of Foreign Language Teaching*, *9*(1), 108–122. https://doi.org/10.5172/ijpl.2011.6.3.260

Pearson, W. S. (2019). Critical perspectives on the IELTS test. *ELT Journal*, *73*(2), 197-206. https://doi.org/10.1093/elt/ccz006

Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European journal of education studies*, *3*(9), 369-387.

Read, J., & Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. In *International English Language Testing System (IELTS) Research Reports 2003: Volume 4* (pp. 153-191). IDP: IELTS Australia Canberra.

Ren, Y. (2011). A study of the washback effects of the College English Test (band 4) on teaching and learning English at tertiary level in China. *International Journal of Pedagogies and Learning*, *6*(3), 243-259. https://doi.org/10.5172/ijpl.2011.6.3.243

Roberts, N., & Thatcher, J. (2009). Conceptualizing and testing formative constructs: Tutorial and

annotated example. *ACM sigmis database: The database for Advances in Information Systems*, *40*(3), 9-39. https://doi.org/10.1145/1592401.1592405

Roopa, S., & Rani, M. S. (2012). Questionnaire designing for a survey. *Journal of Indian Orthodontic Society*, *46*(4), 273-277. https://doi.org/10.5005/jp-journals-10021-1104

Sadler, D. R. (2016). Three in-course assessment reforms to improve higher education learning outcomes. *Assessment & Evaluation in Higher Education*, *41*(7), 1081-1099. https://doi.org/10.1080/02602938.2015.1064858

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least squares structural equation modeling. *Handbook of market research*, *26*(1), 1-40. https://doi.org/10.1007/978-3-319-57413-4_15

Schunk, D. H. (2012). *Learning theories an educational perspective*. Pearson Education, Inc.

Shamsuddin, K., Fadzil, F., Ismail, W. S., Shah, S. A., Omar, K., Muhammad, N. A., Jaffar, A., Ismail, A., & Mahadevan, R. (2013). Correlates of depression, anxiety and stress among Malaysian university students. *Asian J Psychiatr*, *6*(4), 318-323. https://doi.org/10.1016/j.ajp.2013.01.014

Sharma, H., & Aggarwal, A. G. (2019). Finding determinants of e-commerce success: a PLS-SEM approach. *Journal of Advances in Management Research*, *16*(4), 453-471. https://doi.org/10.1108/JAMR-08-2018-0074

Shih, C.-M. (2007). A new washback model of students' learning. *Canadian Modern Language Review*, *64*(1), 135-161. https://doi.org/10.3138/cmlr.64.1.135

Tae-Young, K., & Yoon-Kyoung, K. (2016). A quasi-longitudinal study on English learning motivation and attitudes: The case of South Korean students. *Journal of Asia TEFL*, *13*(2), 72-166. https://doi.org/10.18823/asiatefl.2016.13.2.5.138

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational statistics & data analysis*, *48*(1), 159-205. https://doi.org/10.1016/j.csda.2004.03.005

ThanhNienNews. (2022). *"Cơn sốt" cho con đi học IELTS*. Retrieved 10 May 2024 from https://thanhnien.vn/con-sot-cho-con-di-hoc-ielts-1851506468.htm

Tran, Q. H. (2021). An investigation into non-English major students' problems in taking Aptis listening and reading. *International Journal of TESOL & Education*, *1*(3), 176-191. https://doi.org/10.11250/ijte.01.03.010

Tsagari, D., & Cheng, L. (2017). Washback, Impact, and Consequences Revisited. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language Testing and Assessment* (pp. 359-372). Springer International Publishing. https://doi.org/10.1007/978-3-319-02261-1_24

Tsang, C. L., & Isaacs, T. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing, 39*(2), 212-238. https://doi.org/10.1177/02655322211050600

Wall, D. (2012). Washback in Language Assessment. In *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal1274

Watanabe, Y. (2004). Methodology in washback studies. In *Washback in language testing* (pp. 41-

58). Routledge.

Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS quarterly*, *33*(1), 177-195.

Xiao, W. (2014). The Intensity and Direction of CET Washback on Chinese College Students' Test-taking Strategy Use. *Theory & Practice in Language Studies (TPLS)*, *4*(6), 1171-1177. https://doi.org/10.4304/tpls.4.6.1171-1177

Xiaoju, L. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual & Multicultural Development*, *11*(5), 393-404. https://doi.org/10.1080/01434632.1990.9994425

Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System*, *50*, 56-68. https://doi.org/10.1016/j.system.2015.03.002

Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, *30*(1), 49-70. https://doi.org/10.1177/0265532212442634

Xu, Q., & Liu, J. (2018). Washback: Definitions and Dimentions. In Q. Xu & J. Liu (Eds.), *A Study on the Washback Effects of the Test for English Majors (TEM): Implications for Testing and Teaching Reforms* (pp. 17-22). Springer Singapore. https://doi.org/10.1007/978-981-13-1963-1_2

Zhan, Y., & Wan, Z. H. (2016). Test Takers' Beliefs and Experiences of a High-stakes Computer-based English Listening and Speaking Test. *RELC Journal*, *47*(3), 363-376. https://doi.org/10.1177/0033688216631174

Zhengdong, G. (2009). IELTS Preparation Course and Student IELTS Performance:A Case Study in Hong Kong. *RELC Journal*, *40*(1), 23-41. https://doi.org/10.1177/0033688208101449

Biodata

Nguyen Thanh Minh is a lecturer at Van Lang University, Ho Chi Minh, Vietnam. He has nearly ten years of experience teaching English to both English majors and non-English majors. He is also pursuing his PhD in TESOL at HCMC Open University. His research interests are curriculum design, assessment, predictive modeling, and professional development.