

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

August 2023 – Volume 27, Number 2
<https://doi.org/10.55593/ej.26106a8>

Huu Thanh Minh Nguyen

University of Foreign Language Studies, The University of Danang
<nhtminh@ufl.udn.vn>

Abstract

Understanding the IELTS washback on learning in English as a Foreign Language (EFL) in higher education contexts remains limited. To fill in this research gap, this study investigated washback effects of the International English Language Testing System (hereafter called IELTS) as an English language proficiency exit test for final-year English majors in Central Vietnam from the perspectives of student psychology and learning behaviors by combining Hughes' (1993) and Shih's (2007) washback models to identify the specific washback mechanisms formed by the test factors related to the IELTS. Using a sample of 282 final-year English majors responding to an online questionnaire at a Central Vietnamese University, the factor structures for IELTS washback on learning from test factors were determined by Exploratory Factor Analysis and confirmed by Confirmatory Factor Analysis. Data analyses involved descriptive statistics, Pearson correlation, and multiple linear regression analyses. The study revealed that the IELTS exerted both positive and negative washback on the student psychology and learning behaviors. The students perceived the IELTS as difficult to achieve the required score. They also perceived IELTS as high stakes with various utility values for different aspects of life. The washback mechanisms showed that the test factors exerted effects on the IELTS washback formation in different manners and to different degrees.

Keywords: washback, IELTS, learning behavior, student psychology, final-year English majors

Washback, the effects of tests on teaching and learning, has featured in mainstream language testing research (Hawkey, 2006). Although washback equally pertains to teaching and learning, there has been less washback research on learning than teaching (Wall, 2000; Watanabe, 2004). This is due to the complexity in defining student learning, which can in one definition mean

“student performances, learning motivation, learning strategies, learning practices, or score” (Tsagari, 2007, p. 42). Moreover, washback on learning presents a mixed picture (Shohamy, 2001). Some researchers showed positive washback on learning affecting both learners’ motivation (e.g., Hirai & Koizumi, 2009; Li, 1990; Shohamy, 1993), and target language use and practice for the tests in question (e.g., Allen, 2016; Hung, 2012; Xiao, 2014). However, negative washback on learning was noted through increased learning anxiety (Pan & Newfields, 2012; Shih, 2007), and narrowed learning focused on the test (Damankesh & Babaii, 2015; Ren, 2011; Zhan & Andrews, 2014). Therefore, further washback research on learning is necessary to diversify the literature.

The IELTS, as a high-stakes English language proficiency test (ELPT), is critical to the life chances of test takers because it controls access to international education and employment opportunities. The test is also powerful in reflection of the learners’ identity in the EFL society (Shohamy, 2006). Therefore, IELTS is believed to strongly affect learners (Green, 2006a), with studies about its washback on learning (e.g., Brown, 1998; Elder & O’Loughlin, 2003; Green, 2007a; Mickan & Motteram, 2009; Read & Hayes, 2003) mostly conducted in ESL tertiary-level contexts and for non-English-major students. Little IELTS-related research (e.g., Allen, 2016) has been conducted in EFL higher education settings and on different types of learners and the IELTS washback on learning in Vietnam higher education has not been investigated. Given high-stakes tests such as the IELTS entail washback (Alderson & Wall, 1993) and given washback is a complex phenomenon to be shown in the contexts where tests are used (Green, 2013; Spratt, 2005), the need for more data motivates this study of IELTS washback on the learning of English majors in Vietnam as an under-researched EFL setting.

Research Context

The Vietnamese government has been implementing the National Foreign Language Project that targets comprehensive reforms to foreign language teaching, learning and assessment (Bui & Nguyen, 2016). The project has constituted two prominent achievements in language testing and assessment. One is the development of Vietnam Framework of Foreign Language Competency (VFFLC) to measure the English competence of Vietnamese learners from level 1 to 5, equivalent to A1, A2, B1, B2, C1 in Common European Framework Reference (CEFR). The other is the introduction of Vietnam Standardized Test of English Proficiency (VSTEP) in 2015 as the first national English language test. The VSTEP assesses listening, reading, writing, speaking and is developed in two versions, one measuring levels 1 and 2 and the other measuring levels 3-5 (hereafter called VSTEP.3-5[1]).

Students in Vietnam higher education institutions, under the graduation policy, must take an accredited ELPT of their choice, either domestic (VSTEP.3-5) or international (e.g., IELTS) upon the completion of their study, and obtain a required score to be eligible for degree conferral, in the case of IELTS 7.0. In this sense, at a Central Vietnamese University (CVU, pseudonym) where English majors are trained, English majors can take different ELPTs, and must achieve level 5-VFFLC or C1 level-CEFR. If English majors at the CVU cannot pass their chosen ELPT within two years after program completion, they will fail their degree.

The CVU-delivered English language training program for English majors has been modified in response to the introduction of VSTEP.3-5. English majors since the 2016-2020 cohort, after taking B1- and B2-level courses in the first two years, have studied listening, reading, writing, and speaking in four C1-level modules whose contents are based on the VSTEP.3-5. This major

revision seeks to familiarize students with VSTEP.3-5 and increase its popularity as an English language exit test with far lower test fees than international ELPTs, which have high fees relative to the median income in Vietnam.

Nevertheless, interview evidence from two staff of Academic Training Division (ATD) at the CVU (unpublished), shows that most final-year English majors submitted the IELTS certificate to be screened for graduation eligibility, which proves IELTS's popularity over VSTEP.3-5 and other accepted ELPTs. After the researcher interviewed seven CVU graduates of the cohort 2016-2020, the interview data showed that the IELTS was predominantly chosen for a number of reasons. The graduates referred to VSTEP.3-5, despite being a national test, as only an internal requisite without any use after graduation, because it remains unknown to many employers. Other VSTEP.3-5-related criticisms included the new test format, difficult test content for reading and listening, limited test dates and few test preparation materials. Meanwhile, the reasons in favor of IELTS included its status as popular measure of English proficiency among employers, more available test preparation materials and courses than other ELPTs. The Interviewees also mentioned a measurement of English proficiency was required for the jobs in English-to-Vietnamese/ Vietnamese-to-English translation and interpreting which is what they studied in their major. However, only 356 of 735 English majors of the 2016-2020 cohort could achieve the required passing score and graduate as planned in 2020 (ATD statistics). This means that almost 50% of final-year English majors failed their chosen ELPTs, including the IELTS. In this light, washback of the IELTS might be linked to their failures and need unveiling for final-year English majors to better deal with their learning.

English majors at the CVU mostly come from underdeveloped and low-resourced areas in Central Vietnam, compared with the Southern region where English instruction is more facilitated. Because geographical discrepancies in socio-economic backgrounds determine the quality of English language learning at the expense of poverty-stricken people (Kormos & Kiddle, 2013; Lamb, 2012), students in the Central region are more likely to be at a disadvantage to learn English and pass ELPTs for graduation. More importantly, the high required passing score at C1 level may be linked to pressure in learning (Chen & Squires, 2010; Wang & Liao, 2012), which raises concern whether that benchmark can ensure fairness for the disadvantaged English majors. Taken together, investigating the IELTS washback on English majors in the Central region as a disadvantaged English learning context and exploring the washback mechanisms formed by the factors related to the IELTS may contribute to problematizing the current benchmark for more in-depth insights.

Literature Review

Washback and its models

Various washback models have been proposed to guide washback research, including two developed by Hughes (1993) and by Shih (2007). Hughes (1993) devised a three-part washback model, including participants, processes, and products. Bailey (1996, p. 264) put forward the examples of washback on learning based on Hughes's model, including "practicing items similar in format to those on the test; participating in interactive language practice; reading widely in the target language; listening to non-interactive language; applying test-taking strategies; enrolling in test-preparation courses". These examples refer to the process component of Hughes' model, i.e., "actions taken by the participants which may contribute to the process of learning" (Hughes, 1993, p. 2) as they reflect learning content and strategies that

learners undertake because of a test. In this study, *learning behaviors* including learning content and strategies are defined as the actions learners take because of the design of a specific test, *not* because of learners' more general choice to study English.

Shih (2007) argues that Hughes's model does not fully explain test washback in social and educational contexts where tests are administered because it does not consider individual differences that make washback vary from learner to learner. Therefore, Shih (2007) developed another model viewing washback on learning not only through learning behaviors but also through *student psychology* which, in this study, is concerned with individual differences of students involved in language learning based on their mental experiences, thoughts, and feelings. Shih's model provides an additional description of washback on learning by incorporating the facets of student psychology, i.e., motivation and anxiety because of a test. Thus, combining the process component in Hughes's (1993) model and the facets of student psychology in Shih's (2007) model, this study will frame washback on learning from the perspectives of student psychology and learning behaviors for comprehensiveness in researching washback.

Washback values of high-stakes ELPTs on student psychology and learning behaviors

Washback on learning is both "positive and negative" in its values (Watanabe, 2004, p. 21). If high-stakes ELPTs have positive washback on student psychology, learning motivation is bolstered (Li et al., 2012; Liu & Yu, 2021; Pan, 2009). According to Self-determination theory (Ryan & Deci, 2000), learning motivation can be *intrinsic*, i.e., doing activities for learners' sake to have "a sense of competence" (Oxford, 2013, p. 98), or *extrinsic*, i.e., being provoked by outside needs such as passing the test (Harmer, 2015). L2-English learners' motivation is characterized by various attributes, namely "*motives or reasons* for learning English; personal or academic *goals* in the short or long term; *desire* to achieve those goals; *attitudes* towards English; *interest* in the subject and the learning process; *effort* put into learning in both formal and informal settings; how they *self-regulate* their learning effort over time in the face of distractions and competing goals" (Lamb, 2016, p. 324, original emphasis).

Research has shown several ways tests generate learning motivation based on Lamb's (2016) attributes of L2-English learners' motivation. First, high-stakes ELPTs give students *reasons* to derive the sense of accomplishment from their learning (Dörnyei, 2009; Pan, 2015). Second, because of high-stakes ELPTs, learners set *goals* to change their learning behaviors and achieve the desired test performance (Cavendish et al., 2017). Liu and Yu (2021) found that the Test for English Majors-4 (TEM4) in China gave students extrinsic short-term goals for success in the test and nurtured intrinsic long-term goals to develop English language competences. Third, students adopt positive *attitudes* towards learning English under the policy of high-stakes ELPTs, knowing that it will reflect their English proficiency level (Gan, Humphreys, & Hamp-Lyons, 2004). Fourth, regarding students' learning *interest*, Pan (2009) and Su (2005) showed that students were more captivated with learning English and practicing English skills when driven by high-stakes ELPTs for graduation. Fifth, test takers make considerably more learning *efforts* to both improve English competences (Li et al., 2012; Pan, 2009) and achieve their desired score (Deci & Ryan, 2016) because of high-stakes ELPTs. Fifth, Vietnamese students taking TOEIC for graduation became self-regulated in enhancing their English competences for better employment and education prospects (Nhan, 2013), which reflects other studies in China and Taiwan (e.g., Chu & Yeh, 2017; Shih, 2013; Wu & Lee, 2017). Taken together, high-stakes ELPTs generate intrinsic and extrinsic learning motivation.

In terms of learning behaviors, if tests have positive washback, they promote *holistic learning*, i.e., improving learners' skills and abilities; encouraging emphasis on developing language skills because of the test itself (Alderson & Wall, 1993; Cheng, 1998; Green, 2007a; Gu, 2007). High-stakes ELPTs can promote holistic learning as positive washback in two main ways. First, Liu and Yu (2021) suggested that test takers of TEM4 with intrinsic motivation engaged with reading extensively and listening to authentic materials for English skill development. This finding is nuanced in other studies of the IELTS (Allen, 2016) and College English Test 4 (CET-4) (e.g., Li et al., 2012; Shao, 2006; Sun, 2016; Zhan & Andrews, 2014). Second, students adopt *meta-cognitive strategies* concerned with not only their study strengths but also areas of knowledge and skills for improvements in the actual test (PoPham, 2014; Stecher, 2010).

Negative washback of high-stakes ELPTs on students' psychology involves generating learning anxiety. This is because learners have a fear of poor test performance (Kim & Kim, 2016; Tsagari, 2007). In Confucian cultures, the anxiety is derived from social judgments making students feel anxious about their obligation to succeed in the test and bring glory to their family (Chen et al., 2005; Warden & Lin, 2000). Learners may feel pressure to learn English (Damer & Melendres, 2011; Tsai & Tsou, 2009; Woodrow, 2011). In worst-case scenarios, they may feel too anxious to maintain efforts to learn English (Deci & Ryan, 2016; Ryan & Weistein, 2009), or consider high-stakes ELPTs as burden or even an unachievable goal to pursue (Roderick & Engel, 2001).

One possible effect on learning behaviors of negative washback from high-stakes tests is *restricted learning*, when test takers prioritize such activities as drilling to the test (Damankesh & Babaii, 2015; Dong, 2020; Nguyen, 2020; Tsagari, 2007), and memorizing modeled answers for test tasks (Cheng & Curtis, 2004; Xu & Liu, 2018; Zhan & Andrews, 2014) while other types of learning activities (e.g., extensive reading and out-of-class oral practices) are neglected. High-stakes ELPTs can drive learners to focus on the content tested and desired results rather than engaging with holistic learning (Brown & Abeywickrama, 2010; Chapelle et al., 2011). Gu (2007) and Wang (2007) found that some students spent more time developing *test-wise strategies* rather than developing communicative competences. Under the CET-4 washback, learning content is narrowed to *intense memorization*, e.g., of word lists and sentence patterns and of predicted writing samples in the test (Ren, 2011; Xie & Andrews, 2013). *Test drilling* i.e., narrowing learning content to test-related materials and practices, is considered as restricted learning because it may not be beneficial to develop language proficiency in the target language use domain, as found in studies on the CET-4 washback (Ren, 2011; Xie, 2013, 2015; Zhan & Andrews, 2014). Negative washback was also observed through social strategies: some learners in Mickan and Motteram's (2009) study consulted a knowledgeable mentor on test-wise learning tips to achieve their targeted scores. Interestingly, that finding goes against Karabenick's (2004) observation that high-stakes tests intensify competitiveness among students, thereby limiting seeking external help. However, restricted learning can be culture specific as Asian learners tend to rely more on those surface learning behaviors as a highly valued approach (Liem et al., 2018).

IELTS and its washback on learning

The IELTS can provide language proficiency evidence for candidate selections in different domains (Hamid, 2016). In professional domains, the IELTS is used to decide whether prospective employees can function well in the future workplace (Green, 2019). In education, the IELTS acts as "a global gatekeeping institution" (Pearson, 2019, p. 197) that "measure[s]

English language skills of candidates intending to study in academic or training contexts” (Ingram, 2004, p. 18) and informs the decision of being rejected or accepted to international education institutions based on the prediction of test-takers’ abilities to function in English-medium tertiary academic settings (Thorpe et al., 2017). In terms of its usefulness, the IELTS can enhance students’ employability after graduation, and offer equal learning opportunities in English (Green, 2019), as well as offer a bridge to popular study destinations (e.g., the UK or the USA). The test can provide test takers with “simplified, easy-to-understand, criterion-referenced, and time-bound evidence of English proficiency” (Pearson, 2019, p. 199). However, there are two problems related to the IELTS. First, as part of the lucrative global English language testing industry that is increasingly financially driven (Thorpe et al., 2017), the IELTS has been critiqued from the global economic perspective as an expensive test (Pearson, 2019). “[The] IELTS raise[s] issues of discrimination based on economic inequality, [and] disproportionately impacts on candidates with lower economic means” (Pearson, 2019, p. 201). Relatively high test fees can pressurize test takers to further invest in preparation materials or classroom courses. The second problem is the appropriacy of IELTS as an exit test for graduation. The effect of the IELTS as a matriculation test was believed to be superficial rather than substantial (Humphreys & Gribble, 2013). Crave (2012) found improvements to learning among only some of the students, suggesting that if stakeholders set the target score at 7.0, it might cause difficulty to those standing at level 6.0 in progressing; therefore, learning would plateau.

The extant literature about the IELTS washback on learning as defined in this study only relates to washback on learning behaviors, not student psychology. There is positive washback from IELTS on learning behaviors. According to Allen (2016), Japanese undergraduates studying for the IELTS were more drawn into productive skills practice, and spontaneous speaking activities for daily and abstract topics. Students in Australia adopted *social strategies*, i.e., seeking external help, to boost personal agency and strategic actions in learning for the IELTS (Mickan & Motteram, 2009). Negative washback from IELTS on learning behaviors manifested itself through test drilling and social strategies to seek help with test-wiseness (Mickan & Motteram, 2009). Students preparing for the IELTS tend to do test drilling (Green, 2006b; Mickan & Motteram, 2008), which may not be beneficial in the long term. Another longer term consideration is effective preparation for tertiary study and IELTS preparation has been criticized as not effectively covering certain skills and knowledge needed for academic study in higher education (Dang & Dang, 2021). Those studies, however, do not look at IELTS as an exit test for graduation.

Test factors generating the IELTS washback

Many researchers (e.g., Alderson & Hamp-Lyons, 1996; Brown, 1997; Wall, 1997) attributed washback to numerous factors including contextual factors, learners’ factors, and test factors (Watanabe, 2004). Test factors, in this study, are the attributes of a test that form washback on learning. Given that possible factors related to the IELTS can be linked to the test failure of final-year English majors at the CVU (see “Research context”), investigating the IELTS test factors and how they form washback on the students’ learning is warranted.

Test factors include test methods, test content, skills tested, test purpose, test stakes and test status within the educational system (Watanabe, 2004, p. 22). However, considering later research, Watanabe’s (2004) test factors remain incomplete. First, Shih (2007) put forward test difficulty as one of the components forming washback on learning, with most other factors

overlapping with what is proposed by Watanabe (2004). If learners can anticipate test difficulty based on the communication of construct-related features from test providers, it will improve their learning (Moss, 2016; Jones & Saville, 2006; Saville, 2009). Xie and Andrews (2013) suggest that if test takers accurately perceive the CET-4 difficulty, they will have a higher confidence level, thereby easing learning anxiety. Second, test use forms washback on learning behaviors (Xie & Andrews, 2013) – *test use* refers to social and/or political functions that tests are used to perform (Shohamy, 2001). They concluded that those subscribing to the use of CET-4 as high-stakes and instrumental for extrinsic goals engaged with test-specific behaviors. As test use is a broad term, if the test score helps test takers achieve goals in mind (Xie & Andrews, 2013), that is specifically associated with *test utility* as how useful the test is for test takers to make decisions based on the score interpretation (Bachman, 2005). Therefore, in this study, test use was narrowed down as test utility in forming washback.

There has been limited research (i.e., Stoneman, 2006; Green, 2007b) on how test factors form the IELTS washback on learning. Stoneman (2006) established the relationship between the IELTS stakes and learning motivation in a Hong Kong Polytechnic university. Due to the low stakes (IELTS result not recorded in the transcript) the IELTS did not foster as much learning motivation as other public exams. Green (2007b) investigated how the IELTS test stakes and test difficulty exerted washback on learning. If test takers perceived the IELTS as important and challenging, it would have more intense washback. However, that claim is not specific about whether washback is positive or negative, and which aspects of learning (i.e., student psychology and/or learning behaviors) are subject to washback. Moreover, their studies target only one or two domains of test factors, i.e., test stakes and/or test difficulty. Given the taxonomy of various test factors to date, no research on the IELTS washback on learning have included three test factors: test difficulty, test stakes, test utility and compared which of these factors exert more effects on washback formation to produce more insightful findings.

This Study

The aim of this study was two-fold: 1) explore the washback effects of IELTS on the psychology and learning behaviors of final-year English majors in a Vietnamese-EFL higher education setting as the under-researched subjects, 2) establish a more comprehensive taxonomy of test factors and examine the degree to which they shape the IELTS washback on learning. In this study, test methods, test content, skills tested, test purpose and test status were not investigated because the final-year English majors are aware of the IELTS test status, test design (in relation to its purpose) and the tested skills. These were the driving forces for their investment in learning for IELTS rather than other ELPT alternatives (see “Research context”). This study focuses on how final-year English majors perceive the other test factors: test difficulty, test stakes, test utility, and the degree to which those test factors form IELTS washback on their psychology and learning behaviors.

The study is guided by the following research questions:

- What is the IELTS washback on the psychology and learning behaviors of final-year English majors at the CVU?
- How do final-year English majors at the CVU perceive about the IELTS test factors of test difficulty, test stakes, and test utility?
- Which test factors are more influential on the IELTS washback?

Methodology

General methodological considerations

The study adopted the quantitative approach, addressing questions in causal relationships by generally capturing how IELTS affected learning and how the IELTS test factors accounted for the washback formation. Although it is argued that testing in educational contexts demands the solicited voices of learners as crucial stakeholders (Hamp-Lyons, 2000), washback evidence on learning inferred from limited interviews' accounts (e.g., Allen, 2016) failed the generalizability of findings. As the research aim was not concentrated on eliciting the subjective beliefs that underlay students' reactions because of the IELTS, the quantitative approach using the questionnaire was justified for this study.

Participants

The participants in this study were final-year English majors at the CVU responding to an online questionnaire administered mid-August 2021. The questionnaire was distributed to the online student community of CVU and only accessible via students' institutional email addresses. The responses via the institutional email addresses which indicated the respondents' cohort as final year were retained, which sought to ensure the targeted respondents. There were 285 targeted respondents to the online survey. After the responses having 10% or more missing items and presenting irregular or the same response patterns had been eliminated to have an error-free data set (Curran, 2016), 282 respondents were eligible for further analyses, including 33 males (11.7%) and 249 females (88.3%). Of the valid cases, 44.3% majored in English for Translation and Interpreting ($n = 125$); 30.9% in English for Business ($n = 87$); 24.8% in English for Tourism ($n = 70$). According to the survey result of the students' career plan, they all aspired to secure future employment which is in line with their major in English-to-Vietnamese and Vietnamese-to-English translation and interpreting upon completing their study.

All the participants had yet to take the IELTS and were preparing for it through self-study and extra-curricular preparation courses upon the questionnaire administration. This ensured overt washback, i.e., washback established before the test as the most observable and accessible evidence (Prodromou, 1995). In this sense of overt washback, students' perceptions of the test factors were what the students perceived about the IELTS during the IELTS preparation and before the IELTS test-taking event. The percentage of participants intending to sit the IELTS in September-December 2021, i.e., one to four months from the investigation represented 48.6% ($n = 137$) while 51.4% intended to take it in January-May 2022 ($n = 145$). The end of May 2022 was the institutional deadline to submit IELTS certificates for screening.

Data collection instrument

The study employed a questionnaire which was developed, refined, and piloted (Anderson, 1998) through the three stages as follow.

- (1) Determining operational domains and generating question items

The questionnaire had two main instruments measuring the IELTS washback on learning and test factors forming the washback. The items in the instruments were constructed using a theory-led approach, considering the relevant literature (Lavrakas, 2008). All the items were constructed as 5-point Likert scale, from *1-Strongly Disagree* to *5-Strongly Agree*. The initial pool of items was then scrutinized by five PhD holders in language assessment to examine their face and content validity. This procedure resulted in the inclusion of 30 items for the IELTS

test factors and 27 items for the IELTS washback on learning with justifications from the literature (see Appendices A and B).

(2) Translating and revising

The researcher then produced a Vietnamese version of the questionnaire which was proofread by two lecturers instructing translation modules at the CVU and three non-expert native speakers of Vietnamese to check if the items were comprehensible.

(3) Piloting

The revised questionnaire went through a pilot study on non-targeted participants. This was done through cognitive interviewing (Jobe & Herrmann, 1996) where the researcher queried seven piloted participants whether their comprehension of the questions was aligned with the researcher's intent. Following the interviews, incomprehensible wording caused by English-Vietnamese translation, and the misunderstood items were revised. Twelve other students completed the revised questionnaire as a second piloting to further check for clarity and estimate the time for completion.

Instrumentation

To determine the factor structure and confirm the construct validity of the two instruments, the total sample was randomly split, with the first half used for Exploratory Factor Analyses (EFA; $n = 141$) and the second half for Confirmatory Factor Analysis (CFA; $n = 141$) (Lorenzo-Seva, 2022). The EFA were conducted in SPSS 23 on the first half of the sample by principal component analysis and Varimax rotation. To explore the number of factors, Kaiser's eigenvalues-greater-than-one criterion was employed (Kaiser, 1960). Items showing loadings less than 0.4 (DeVellis, 2003) and/or cross-loading on two or more factors with loadings of 0.4 or greater (Krishnan, 2011) were excluded. Following the EFA from which the factor structure was derived, the CFA were then undertaken on the second half by AMOS 22 to examine the model fit of the items into latent constructs of the IELTS washback and test factors. The values of chi-square (χ^2), degrees of freedom (df), and other fit indices, i.e., the comparative fit index (CFI); the Tucker-Lewis Index (TLI); the goodness of fit index (GFI), the root mean square error of approximation (RMSEA) should be reported in that $\chi^2/df \leq 2$; CFI, TLI and GFI ≥ 0.9 ; RMSEA ≤ 0.08 are considered good fits (Hair et al., 2010).

Following the procedure above, the EFA of the IELTS test factor instrument generated three factors with 13 items and cumulatively explained 57.497% of the total variance, thereby fitting very well the intended structure of the instrument (Appendix C). The measurement model for the test factors (Figure 1) satisfactorily fits the data: $\chi^2 = 96.494$, $df = 62$, $p = .000$, $\chi^2/df = 1.556$, CFI = .935, TLI = .918, GFI = .921, RMSEA = .057, which explained the underlying pattern of the latent constructs. The EFA of the IELTS washback on learning instrument generated four factors with 13 items and cumulatively explained 64.625% of the total variance, thereby fitting very well the intended structure of the instrument (Appendix D). The measurement model for the IELTS washback (Figure 2) has the satisfactory fit indices: $\chi^2 = 113.659$, $df = 59$, $p = .000$, $\chi^2/df = 1.926$, CFI = .924, TLI = .900, GFI = .909, RMSEA = .074, which explained the underlying pattern of the latent constructs.

The test factors instrument measured three constructs, namely test difficulty ($\alpha = 0.70/ 4$ items, e.g., *I think the required passing score on the IELTS will be difficult to achieve*), test stakes ($\alpha = 0.72/ 5$ items, e.g., *The IELTS result decides if I can graduate on time*), test utility ($\alpha = 0.74/$

4 items, e.g., *The IELTS results helps me to gain employment after graduation*). The IELTS washback on learning measured four constructs, namely learning motivation ($\alpha = 0.74/ 3$ items, e.g., *Because of the IELTS, I set goals to develop English language skills*), learning anxiety ($\alpha = 0.70/ 3$ items, e.g., *Because of the IELTS, I am under pressure to learn English*), holistic learning ($\alpha = 0.76/ 3$ items, e.g., *Having to take the IELTS makes me deal with gaps in my English language competence*), restricted learning ($\alpha = 0.76/ 4$ items, e.g., *I memorize sample answers to the questions that are likely to appear in the real test*).

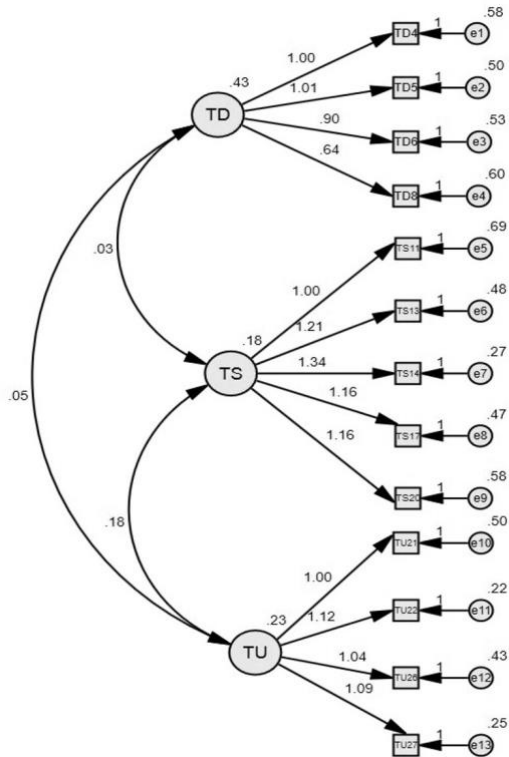


Figure 1. Measurement model for the test factors.

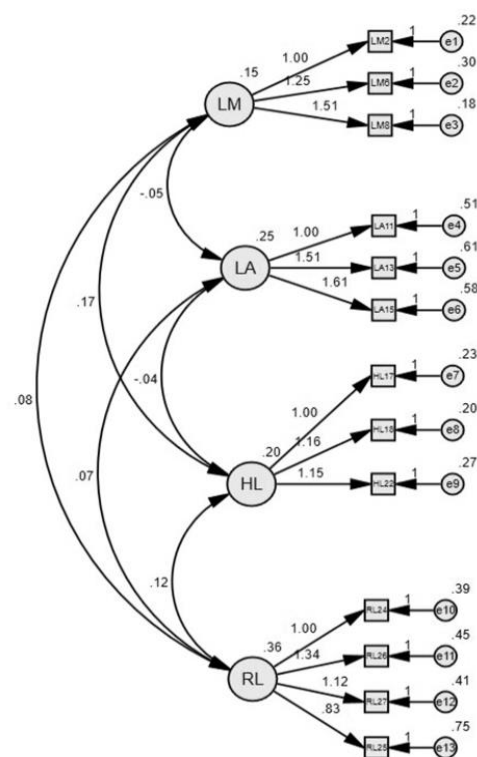


Figure 2. Measurement model for the IELTS washback on learning.

Data analyses

On the whole sample ($n = 282$), all the items were processed in SPSS 23. Descriptive statistics were first generated to unfold students' perceptions of the IELTS washback on learning and test factors. Pearson correlation (r) values were then generated to demonstrate how the test factors were associated with the IELTS washback on learning. Based on the established correlations, multiple linear regression analyses were subsequently used to show which test factors had more effect in forming the IELTS washback on learning via comparison of standardized beta coefficients (β). To achieve non-collinearity in each of the regression models, the VIF values should be under 2.0, and the condition indexes should be lower than 30 (Kim, 2019). Moreover, the variance proportions of all independent variables should not be larger than 0.9 (Hair et al., 2010).

Ethics

The ethical considerations included informed consents of participants, protected participants' identity, as well as commitment to using the collected data only for the research and not for determining the participants' grades. The online questionnaire in Vietnamese administered to the participants had a cover page, with an information sheet stating the research aim and covering the above ethical considerations, as well as a consent form signed by the participants (Mann, 1994).

Findings

Perceptions of the IELTS washback on learning

According to Table 1, the IELTS had positive washback on student psychology. The students were intrinsically motivated, with over 90% and 83% setting the goal and making effort outside English classes respectively for developing English skills. Moreover, over 85% were extrinsically motivated to achieve the targeted score. Nevertheless, the IELTS also had negative washback on student psychology, with over 70% feeling pressure to learn English. However, the IELTS did not drive the students to feel discouraged from learning English as over 52% disagreed and totally disagreed with this perception.

The IELTS had positive washback on learning behaviors. Because of the IELTS, almost 80% of the students worked on areas for improvement in their English language competence as metacognitive strategies, focused on developing English skills, and adopted social strategies to seek external help from their tutor under the IELTS washback context. However, negative washback of the IELTS on learning behaviors was also evident. Drilling to the test, learning test-taking tips, memorizing sample answers, using social strategies for test-wise learning tips were undertaken by most of the student respondents.

Perceptions of the test factors

According to Table 2, most of the students found the IELTS difficult in achieving the required score and dealing with test questions. Regarding test stakes, almost all students considered the IELTS to be decisive to their graduation. The majority concurred on the IELTS as being a high stakes test for their major-aligned and desired employment as well as their choice of education institutions overseas in the future. They also found that the IELTS result determined social judgments of their outcomes at the CVU.

Students also perceived the IELTS result with various utility values, in terms of long-term goals for future employment and learning. They also found the IELTS result instrumental in enhancing competitiveness for seeking and securing jobs as well as in shaping their future trajectories in learning English. Notably, over 85% ascribed gaining equal access to learning opportunities to their IELTS result.

Table 1. The IELTS washback on learning

	Statements	Strongly Disagree (%)	Disagree (%)	Unsure (%)	Agree (%)	Strongly Agree (%)
LM	LM2. Because of the IELTS, I set goals to develop English skills.	0.0	0.6	8.2	60.0	31.2
	LM6. Because of the IELTS, I make effort to improve English skills outside English classes.	0.6	1.2	15.3	56.5	26.5
	LM8. I set goals to achieve the targeted test score.	0.6	1.8	11.2	55.3	31.2
LA	LA11. Because of the IELTS, I feel pressure to learn English.	0.6	10.6	20.6	52.9	15.3
	LA13. Because of the IELTS, I feel discouraged from learning English.	14.1	37.9	28.2	14.5	5.3
	LA15. I have a lot of self-doubt about my ability to pass the IELTS.	8.2	15.9	32.4	31.8	11.8
HL	HL17. Because of the IELTS, I spend lots of time developing English skills.	0.0	2.4	19.4	63.5	14.7
	HL18. Having to take the IELTS makes me aware of weaknesses in my English language competence.	0.0	2.4	18.8	60.6	18.2
	HL22. Because of the IELTS, I seek help from my tutor to guide my own learning.	1.2	1.2	20.6	59.4	17.6
RL	RL24. My learning is narrowed to practicing mock tests to achieve the targeted score.	1.8	8.2	24.7	52.4	12.9
	RL26. I memorize sample answers to the questions that are likely to appear in the real test.	7.6	12.5	21.8	39.9	18.2
	RL27. My learning is narrowed to learning test-taking tips to achieve the targeted score.	4.7	8.9	27.6	48.0	10.8
	RL25. I consult my tutor only on test-taking tips to achieve the targeted score.	2.4	12.9	27.1	47.1	10.6

Table 2. Perceived test factors

	Statements	Strongly Disagree (%)	Disagree (%)	Unsure (%)	Agree (%)	Strongly Agree (%)
TD	TD4. I think that the required passing score on the IELTS will be difficult to achieve.	2.9	8.0	30.9	42.3	15.9
	TD5. I think that the IELTS is more difficult than the other test alternatives.	2.4	18.8	40.6	29.4	8.8
	TD6. The time allotted to complete the IELTS is too short.	2.4	12.9	40.6	33.9	10.2
	TD8. I think that the questions in the IELTS test are difficult to answer.	2.4	6.1	32.9	39.2	19.4
TS	TS11. The IELTS result decides if I can graduate on time.	0.0	0.0	3.2	63.5	33.3
	TS13. The IELTS result affects my employability for desired jobs that involve using language skills.	1.2	4.1	18.8	47.1	28.8
	TS14. The IELTS result shows whether I am qualified to work in an area that fits my university major.	1.2	1.2	18.8	52.4	26.5
	TS17. The IELTS result affects my choice for higher education institutions abroad.	1.2	4.7	17.6	51.8	24.7
	TS20. The IELTS result affects how people judge the outcomes of my time at university.	1.8	7.1	24.7	46.5	20.0
TU	TU21. The IELTS result helps me to gain employment after graduation.	2.9	2.9	16.7	59.5	18.0
	TU22. The IELTS result helps me to enhance my competitiveness in future job seeking.	1.2	0.6	8.8	57.1	32.4
	TU26. The IELTS result helps me to be clearer about where I should go next in English learning.	0.6	5.3	17.6	52.9	23.5
	TU27. The IELTS result helps me to gain equal access to learning opportunities.	0.6	2.4	11.2	58.8	27.1

Effects of test factors on the IELTS washback

Test factors had several correlations with the IELTS washback (Table 3). Test difficulty was significantly correlated with the negative washback variables, i.e., learning anxiety and restricted learning. Meanwhile, it had no significant correlation with the positive washback variables, i.e., learning motivation and holistic learning. Test stakes, while being associated with learning anxiety, had no correlation with learning motivation. In terms of learning

behaviors, test stakes showed significant correlations with both holistic and restricted learning. Test utility significantly correlated with all the IELTS washback variables, both positive and negative.

Table 3. Correlations between test factors and the IELTS washback

		Test difficulty	Test stakes	Test utility
Learning motivation	r	-.06	.14	.24**
	Sig.	.414	.068	.001
	95% Confidence Interval	[-.181 - .066]	[.069 - .349]	[.107 - .390]
Learning anxiety	r	.54**	.21**	.20**
	Sig.	.000	.005	.009
	95% Confidence Interval	[.482 - .784]	[.088 - .326]	[.070 - .485]
Holistic learning	r	-.07	.31**	.29**
	Sig.	.399	.000	.000
	95% Confidence Interval	[-.178 - .071]	[.161 - .435]	[.140 - .422]
Restricted learning	r	.26**	.24**	.21**
	Sig.	.001	.000	.006
	95% Confidence Interval	[.117 - .431]	[.108 - .474]	[.075 - .451]

**Correlation is significant at 0.05 level (2-tailed).

To predict the effects of test factors on the IELTS washback on learning, multiple linear regression analyses were conducted. According to Table 4 and 5, the data suggested non-collinearity between dependent and independent variables, thereby separating linear regression analyses of each model (Hair et al., 2010; Kim, 2019).

Table 4. Collinearity statistics

Model		Collinearity statistics	
		Tolerance	VIF
1	(Constant)		
	Test difficulty	.986	1.014
	Test stakes	.574	1.742
	Test utility	.569	1.758

Table 5. Collinearity diagnostics

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				Constant	Test difficulty	Test stakes	Test utility
1	1	3.947	1.000	.00	.00	.00	.00
	2	.036	10.520	.00	.79	.06	.04
	3	.010	19.591	.99	.20	.17	.08
	4	.007	23.034	.01	.00	.77	.88

Linear regression analyses showed that the test factors had effects on the IELTS washback formation in different manners and to different degrees (Table 6). All the test factors were not predictors of learning motivation, but rather generated learning anxiety. Specifically, test difficulty had the greatest effect on learning anxiety ($\beta = .522, p = .000$). Test utility had greater effects on learning anxiety than test stakes ($\beta = .24, p = .002$ and $\beta = .17, p = .029$ respectively). All the test factors accounted for 29.6% of the variance in learning anxiety as the IELTS washback.

As regards the linear regression models on holistic and restricted learning, test stakes and test utility formed holistic learning, with test stakes having greater effects than test utility ($\beta = .31$ and $\beta = .23$ respectively, $p = 0.000$). Only test difficulty generated restricted learning ($\beta = .19, p = .000$).

Table 6. Effects of the test factors on the IELTS washback

Dependent Variable	Independent Variable	Adjusted R ²	Overall p-value	β	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Learning motivation	Test difficulty			-.01	.906	-.204	.037
	Test stakes	.065	.003	.04	.668	.015	.269
	Test utility			.10	.338	.096	.390
Learning anxiety	Test difficulty			.522	.000	.463	.765
	Test stakes	.296	.000	.17	.029	.064	.305
	Test utility			.24	.002	.069	.406
Holistic learning	Test difficulty			-.09	.346	-.200	.037
	Test stakes	.105	.000	.31	.005	.029	.409
	Test utility			.23	.037	.026	.385
Restricted learning	Test difficulty			.19	.008	.097	.452
	Test stakes	.100	.000	.10	.225	.086	.427
	Test utility			.06	.479	.048	.407

Discussion

The IELTS washback on student psychology and learning behaviors

Regarding IELTS washback on students' psychology, the IELTS generated intrinsic learning motivation for setting the goal and making effort outside English classes to develop English skills. Extrinsic learning motivation for achieving the targeted score because of the IELTS was also found. These findings differ from previous studies (e.g., Deci & Ryan, 2016; Li et al., 2012; Pan, 2009) in that learners in those studies, as non-English majors, had either intrinsic motivation to learn English or extrinsic motivation because of the test. The findings are, however, in line with Liu and Yu's (2021) finding that the TEM4 motivated English-major test takers intrinsically to develop English competence and extrinsically to succeed in the test. It is implied that English majors generate more forms of motivation for learning English under the

test washback than non-English majors. Irrespective of the motivations, there remained instances of negative washback from pressure to learn English on the students' psychology, which corroborates previous studies (e.g., Damer & Melendres, 2011; Tsai & Tsou, 2009; Woodrow, 2011). However, learning anxiety was not far-reaching to such severity that the students could have felt discouraged from learning, which is inconsistent with Pan and Newfields (2012) and Roderick and Engel (2001). This is because those studies investigated low-level learners who were more likely to abandon learning efforts because of high-stakes test pressure. Students in this study passed the C1-level modules in their previous English courses and may have devised strategies to sustain learning for the IELTS as a result.

Turning now to positive washback on learning behaviors, because of the IELTS, metacognitive strategies were adopted for improvements in learning English, which echoes Stecher's (2010) and PoPham's (2014) findings. The use of social strategies concurs with the learners' behaviors in Mickan and Motteram (2009). Focused engagement with developing English skills because of the IELTS is also in concert with Allen's (2016) finding. This might be due to their intrinsic motivation for improving English skills (Li et al, 2012; Liu & Yu, 2021; Shao, 2006). However, the relationship between intrinsic learning motivation and holistic learning behaviors because of the IELTS should be further examined.

Nevertheless, drilling to the test, learning test-taking tips, and memorizing sample answers were also employed at the same time as negative washback, which concurs with previous investigations into CET-4 (e.g., Gu, 2005; Ren, 2011; Xie, 2013). The students also sought test-wise learning tips from social strategies, which reflects Mickan and Motteram's (2009) finding, but is not consistent with Karabenick's (2004) view about the limited impact of social strategies for high-stakes tests. This is because the IELTS in the CVU does not involve the students' competitiveness through classifying their academic standing, but rather only decides whether students are eligible to be awarded a degree. The negative washback on learning behaviors could be for three reasons. First, because almost half of the participants intended to take the IELTS only one-four months away from the study, they might have shown purposefully restricted preparation to maximize their learning as the test was approaching (Xie & Andrews, 2013). Second, the students' extrinsic motivation to perform well on the IELTS might have led to increased focus on memorization and test wiseness (Zhan & Wan, 2014). However, the relationship between extrinsic learning motivation and restricted learning behaviours in the IELTS washback context requires further investigations. Third, a group of students in the study might have had their current standing of English proficiency which was equivalent to IELTS 6.0 and faced difficulty progressing to IELTS 7.0 as the required score for graduation (Craven, 2012). Because of this plateau in learning as suggested by Craven (2012), the students might have restricted their learning only for the desired result without concern about whether holistic learning might have occurred.

On balance, both positive and negative washback of the IELTS on learning behaviors resonate with typical characteristics of Asian learners who are stereotyped as simultaneously adopting both holistic and restricted learning as their valued learning style (Liem et al., 2008). The students concurrently trying to develop their English skills and obtain high score under this IELTS washback context could derive certain value from adopting holistic and restricted learning simultaneously in achieving their goals, which strengthens Shohamy's (2001) stance that a reputed test does not necessarily result in positive washback, and that negative washback is not always bad.

Perceptions of test factors

The IELTS was perceived as difficult in achieving the required score and dealing with test questions. This could be because instructional content at the CVU follows the format and question types of the VSTEP.3-5 which has little in common with the IELTS. The perceived test difficulty was therefore attributed to the discrepancy between what is instructed and what is tested. For the economically disadvantaged students, feelings of being challenged by a high required passing score on the IELTS might have been due to their insufficient economic resources to develop English proficiency through the expensive IELTS test preparation process (Pearson, 2019) and pay high test fees relative to their income.

As for test stakes, the IELTS was considered decisive to the students' graduation, which reflects their understanding of its importance to their learning as the degree requirement. The IELTS is high stakes for the students' employment and education prospects, which reflects Green's (2019) and Pearson's (2019) universal views about the IELTS in deciding one's qualification for employment and education. Moreover, the IELTS stakes in deciding social judgments of the students' outcomes at CVU can reflect Shohamy's (2006) view that success in high-stakes ELPTs is a valued commodity for those living in a society where English is not a major language.

The students' perceptions of the IELTS result with various utility values to achieve long-term goals for future employment and learning (Green, 2019; Pearson, 2019). Notably, as the students in this study from socio-economically disadvantaged groups might not be able to claim educational achievements, proving English competences through the IELTS will afford them equal access to future learning opportunities that knowledge of English opens up, although not all can be beneficiaries of the IELTS in the same manner.

The IELTS washback mechanisms formed by the test factors

In terms of washback mechanisms on student psychology, all the test factors did not shape learning motivation, but rather aroused learning anxiety as negative washback. Learning anxiety resulting from the high stakes could be explained by high anxiety levels felt by students about high-stakes tests (Kim & Kim, 2016) because passing them with good scores affects future employment and educational opportunities. Another possible reason is Confucian collectivism in Vietnam that could have made the students feel anxious about their obligation to succeed in the test and bring glory to themselves and their family (Chen et al., 2005; Warden & Lin, 2000). Although this finding from the test stakes perspective somewhat corroborated Stoneman's (2006) study about the IELTS generating little learning motivation, the consistency was superficial. Students in Stoneman's study were in Hong Kong where English is commonly used and the IELTS result was not recorded in their transcript and would not affect their future opportunities in the same way as the students in this EFL context whose degree would be withheld without an IELTS 7.0 (see "Research Context") and whose future employment and education are partly decided by the IELTS result (Green, 2019; Thorpe et al., 2017). Therefore, the high-stakes nature of IELTS in this research explained the generation of learning anxiety over learning motivation, as opposed to limited motivation explained by the IELTS low stake in Stoneman's study.

Learning anxiety caused by test difficulty, however, was not comparable to Xie and Andrews (2013) observed relationship between students' understanding of CET-4 test demands and higher confidence level. That is because their study differently elicited students' perceptions of

what is required to perform question types, rather than the perceived difficulty of test questions and in required score, as in this study. Yet, this finding may extend Green's (2007b) generalized view of the IELTS washback on learning in that test difficulty is a driver of negative washback on student psychology. Moreover, of all the test factors, test difficulty had the greatest effect on generating learning anxiety, probably due to the added challenge in achieving the required score. The explained multitude of effects may elaborate Green's (2007b) view of IELTS washback intensity being explained by test difficulty and test stakes, with the former having a greater effect than the latter. However, both developments of Green's view should be cautioned, considered in context.

Comparing test stakes and test utility, perceiving the IELTS to have various utility values caused greater learning anxiety than perceiving it as high stakes. This is likely because the students were more worried about how the IELTS result and their learning could be useful for achieving long-term goals of employment and education than about the test stakes, because they would be still given two more years after program completion to pass the test.

In terms of washback mechanisms on learning behaviors, both test stakes and test utility only resulted in holistic learning. This finding might be traceable to Nhan (2013) showing Vietnamese students' autonomy in learning English for TOEIC as a high-stakes ELPT for better employment and educational prospects. That might be comparable to the students in this study seeking to enhance English competences for occupational and educational opportunities through IELTS. However, this finding is not aligned with Xie and Andrews's (2013) study about CET-4's perceptions about high stakes tests and test utility as an instrument for achieving extrinsic goals, leading to test-specific learning (e.g., memorizing, test drilling). One possible reason is that their study was based in a lower-tier university where students tended to demonstrate lower English ability and were more likely to spend more time and resources on test-specific preparation to avoid test failure (Xie, 2015). Another possible reason is that students in this study took the IELTS both to meet graduation eligibility and as a means of securing future employment and education.

Comparing the effects of test stakes and test utility, test stakes had greater effect on forming holistic learning than test utility. This might be because the IELTS was perceived as decisive to how satisfactory students' language abilities are for graduation, and their eligibility for the jobs fitting their major (e.g., those in the industries involving using four language skills in totality such as translation and interpreting). In the sense of eligibility for future plans, the stakes count more in forming holistic learning under IELTS washback than the utility, which is limited to employment and learning opportunities in general.

Test difficulty was the only factor leading to restricted learning. Because the IELTS is a different test from VSTEP which CVU students were prepared for and their formal English curriculum is based on, transitioning to studying for the IELTS may have caused difficulty to the students and made them narrow their learning to be test specific. This finding could extend Green's (2007b) view that if a test is thought to be challenging, negative washback is exerted on learning behaviors.

Conclusions

This study explored how the IELTS as an exit test affected the psychology and learning behaviors of final-year English majors and the washback mechanisms formed by the test factors. What underscored the findings is both positive and negative washback of the IELTS on

the student psychology and learning behaviors at the same time. The students generally perceived the IELTS as presenting difficulty in answering test questions and achieving the required score, and as a high stakes test for graduation, future employment, future education, and social judgments. IELTS was also found to have various utility values for future employment and learning. All the test factors did not shape learning motivation, but rather caused learning anxiety. Test difficulty had the greatest effect while test utility exerted greater effect than test stakes on generating learning anxiety. Both test stakes and test utility contributed to generating holistic learning, with the test stakes exerting greater effects than the test utility. Only test difficulty made students restrict learning.

Some implications from the findings: students had both intrinsic and extrinsic motivation to achieve the targeted score, which easily leads to restricted learning (Cheng & Curtis, 2004). Students should be exposed to authentic learning materials through which they should attempt to develop English proficiency to retain long-term intrinsic motivation, given that the test is a short-term extrinsic motivator. Because IELTS is perceived as high stakes with utility values contributing to holistic learning, students should always regulate their learning towards how the test decides their life chances in tertiary studies and in future intentions. Lastly, because the target score as band 7.0 may not ensure learning gains for those at low proficiency levels and even involves difficulty that may restrict learning, the students should be aware of their current standing to decide whether they should use the IELTS for exit purposes in order not to sacrifice holistic learning. Teachers should change students' beliefs in intense memorization because of the IELTS. Although that can only be collectively achieved via the educational system, classroom teaching in EFL contexts can affect students' learning (Dong, 2020). Teachers may teach alternative learning strategies from which students know how to expand their language repertoires for learning. One reason why students do intense memorization is that this is the only strategy they are taught (Li & Cutting, 2011). Second, given that restricted learning might stem from perceiving the IELTS as difficult because it is a different test from what curriculum instruction is designed for, lecturers need to assure students that what they learn in classroom settings is transferrable as a preparation for any test. This can be done by enhancing classroom provisions towards diversifying communicative teaching approaches and enriching meaning-focused teaching contents for students to continuously improve their proficiency level and deal with any test. Third, providing emotional support and helping students discern their learning progress through constructive feedback mechanisms is necessary to compensate for learning anxiety resulting from high test stakes and utility values.

This study has certain limitations. First, only a questionnaire was employed, thereby not reflecting the in-depth IELTS washback data. Given the role of interviews (Muñoz & Álvarez, 2010) and classroom observations (Morrow, 1986) in washback research to determine whether learners' perceptions are reflected in their behaviors, further studies may combine questionnaires with interviews or classroom observations for triangulation. Second, although the findings can be generalized to the English majors at the CVU, caution should be exercised in generalizing the findings to English majors in Northern and Southern regions of Vietnam. Future studies may involve participants from the other regions and investigate differential washback across regions and disciplines as English-majored or non-English majored. Despite the limitations, the study documented the first attempt to investigate the IELTS washback on both student psychology and learning behaviors and its washback mechanisms formed by the test factors in EFL tertiary-level contexts and to English majors as under-researched subjects.

About the author

Huu Thanh Minh Nguyen is a lecturer in English as a Foreign Language at University of Foreign Language Studies, The University of Danang. He earned his Master's degree in TESOL Studies from the University of Leeds, UK. His main research interests include Language Testing and Assessment, Second Language Writing, and Corpus Linguistics. ORCID ID: 0000-0002-9494-9641.

To Cite this Article

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. *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 27(2). <https://doi.org/10.55593/ej.27106a8>

Acknowledgements

I would like to greatly thank Associate Professor Dr. Jason Fan, Deputy Director at the Language Testing Research Center, University of Melbourne, the editor, and anonymous reviewers whose insightful and critical reviews have substantially contributed to improving the earlier versions of the manuscript. Special thanks are also extended to all the participants without whom the study would not have been materialized.

Note

[1] Information on the VSTEP.3-5 test format including skills tested and question types can be found via the link: <http://vstep.vn/test-format/>.

References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13, 280–297. <https://doi.org/10.1177/026553229601300304>
- 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), 1-20. <https://doi.org/10.1186/s40468-016-0030-z>
- Anderson, G. (1998). *Fundamentals of Educational Research* (2nd ed.). Taylor & Francis.
- Bachman, L. F. (2005). Building and Supporting a Case for Test Use. *Language Assessment Quarterly*, 2(1), 1-34. https://doi.org/10.1207/s15434311laq0201_1
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257–278. <https://doi.org/10.1177/026553229601300303>
- Brown, J. D. (1997). Do tests washback on the language classroom? *The TESOLANZ Journal*, 5(5), 63-80.
- Brown, J. D. (1998). An investigation into approaches to IELTS preparation, with particular focus on the Academic Writing component of the test. In S. Wood (Ed.) *IELTS Research Reports* (Vol. 1, pp. 20-37). ELICOS/IELTS Australia.

- Brown, H. D., & Abeywickrama, P. (2010). *Language Assessment: Principles and classroom practices*. Pearson Education.
- 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 *English language education policy in Asia* (pp. 363-388). Springer.
- Cavendish, W., Márquez, A., Roberts, M., Suarez, K., & Lima, W. (2017). Student engagement in high-stakes accountability systems. *Penn GSE Perspectives on Urban Education*, 14(1), 1-3.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2011). *Building a validity argument for the test of English as a Foreign Language*. Routledge.
- Chen, M. L., & Squires, D. (2010). Vocational college students' perceptions on standardized English proficiency tests. *The Asian EFL Journal*, 12(2), 68-91.
- Chen, J. F., Warden, C. A., & Chang, H. (2005). Motivators that do not motivate: The case of Chinese EFL learners and the influence of culture on motivation. *TESOL Quarterly*, 39(4), 609-633. <https://doi.org/10.2307/3588524>
- 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](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>
- 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>
- Craven, E. (2012). The Quest for IELTS 7.0: Investigating English Language Proficiency of International Students in Australian Universities. *IELTS Research Reports*, 13, 1-61.
- 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>
- Damer, D. E., & Melendres, L. T. (2011). "Tackling test anxiety": A group for college students. *The Journal for Specialists in Group Work*, 36(3), 163-177. <https://doi.org/10.1080/01933922.2011.586016>
- Dang, C. N., & Dang, T. N. Y. (2021). The Predictive Validity of the IELTS Test and Contribution of IELTS Preparation Courses to International Students' Subsequent Academic Study: Insights from Vietnamese International Students in the UK. *RELC Journal*, 54(1), 84-98. <https://doi.org/10.1177/0033688220985533>
- Deci, E. L., & Ryan, R. M. (2016). Optimizing Students' Motivation in the Era of Testing and Pressure: A Self-Determination Theory Perspective. In W. L. J. Wang, & R. Ryan (Eds.). *Building Autonomous Learners* (pp. 9-29). Springer.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Sage Publications.

- Dong, M. (2020). Structural relationships 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, 1-11. <https://doi.org/10.1016/j.stueduc.2019.100824>
- Dörnyei, Z. (2009). The L2 motivational self-system. *Motivation, language identity and the L2 self*, 36(3), 9-11. <https://doi.org/10.21832/9781847691293-003>
- Elder, C., & O' Loughlin, K. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS. In R. Tulloh (Ed.) *IELTS Research Reports* (Vol. 4, pp. 207-254). IELTS Australia.
- Gan, Z., Humphreys, G., & Hamp-Lyons, L. (2004). Understanding successful and unsuccessful EFL students in Chinese universities. *The Modern Language Journal*, 88(2), 229-244. <https://doi.org/10.1111/j.0026-7902.2004.00227.x>
- Green, A. (2006a). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes, *Assessing Writing*, 11(2), 113-134. <https://doi.org/10.1016/j.asw.2006.07.002>
- Green, A. (2006b). Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly*, 3(4), 333-368. <https://doi.org/10.1080/15434300701333152>
- Green, A. (2007a). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-session language courses. *Assessment in Education*, 14(1), 75-97. <https://doi.org/10.1080/09695940701272880>
- Green, A. (2007b). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge University Press.
- 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>
- Green, A. (2019). Restoring perspective on the IELTS test. *ELT Journal*, 73(2), 207-215. <https://doi.org/10.1093/elt/ccz008>
- Gu, X. (2007). *Positive or negative? – An empirical study of CET washback*. Chongqing University Press.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson.
- Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. *System*, 28, 579-591. [https://doi.org/10.1016/S0346-251X\(00\)00039-7](https://doi.org/10.1016/S0346-251X(00)00039-7)
- Hamid, M. O. (2016). Policies of global English tests: test-takers' perspectives on the IELTS retake policy. *Discourse: Studies in the Cultural Politics of Education*, 37(3), 472-87. <https://doi.org/10.1080/01596306.2015.1061978>
- Harmer, J. (2015). *The Practice of English Language Teaching* (5th ed.). Pearson Education Limited.
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge University Press.

- 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.
- Humphreys, P., & Gribble, C. (2013). *English language and the transition to further study*. International Education Association Australia. <http://www.ieaa.org.au/documents/item/54>
- 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>
- Ingram, D. (2004). Towards more authenticity in language testing. *Babel*, 39(2), 16-24.
- Jobe, J. B., & Herrmann, D. J. (1996). Implications of Models of Survey Cognition for Memory Theory. In D. J. Herrmann, C. McEvoy, C. Herzog, P. Hertel, & M. K. Johnson (Eds.) *Basic and Applied Memory Research: Practical Applications* (pp. 193-205). Erlbaum.
- Jones, N., & Saville, N. (2016). *Learning Oriented Assessment: A systemic approach*. UCLES & Cambridge University Press.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151. <https://doi.org/10.1177/001316446002000116>
- Karabenick, S. A. (2004). Perceived achievement goal structure and college student help seeking. *Journal of Educational Psychology*, 96(3), 569-581. <https://doi.org/10.1037/0022-0663.96.3.569>
- Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 72(6), 558-569. <https://doi.org/10.4097/kja.19087>
- Kim, T. Y. (2010). Socio-political influences on EFL motivation and attitudes: Comparative surveys of Korean high school students. *Asia Pacific Education Review*, 11(2), 211-222. <https://doi.org/10.1007/s12564-010-9071-7>
- Kim, T. Y., & Kim, Y. K. (2016). A quasi-longitudinal study on English learning motivation and attitudes: The case of South Korean students. *The Journal of Asia TEFL*, 13(2), 138-155. <https://doi.org/10.18823/asiatefl.2016.13.2.5.138>
- Kormos, J., & Kiddle, T. (2013). The role of socio-economic factors in motivation to learn English as a foreign language: The case of Chile. *System*, 41(2), 399-412. <https://doi.org/10.1016/j.system.2013.03.006>
- Lamb, M. (2012). A self-system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning*, 62(4), 997-1023. <https://doi.org/10.1111/j.1467-9922.2012.00719.x>
- Lamb, M. (2016). Motivation. In G. Hall (Ed.) *The Routledge Handbook of English Language Teaching* (pp. 324-337). Routledge.
- Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Sage Publications.

- Li, X. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual and Multicultural Development*, 11(5), 393-404. <https://doi.org/10.1080/01434632.1990.9994425>
- Li, H., Zhong, Q., & Suen, H. K. (2012). Students' Perceptions of the Impact of the College English Test. *Language Testing in Asia*, 2, 77-94. <https://doi.org/10.1186/2229-0443-2-3-77>
- Liem, A. D., Lau, S., & Nie, Y. (2008). The role of self-efficacy, task value, and achievement goals in predicting learning strategies, task disengagement, peer relationship, and achievement outcome. *Contemporary Educational Psychology*, 33(4), 486–512. <https://doi.org/10.1016/j.cedpsych.2007.08.001>
- 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, 1-10. <https://doi.org/10.1016/j.stueduc.2020.100967>
- Lorenzo-Seva, U. (2022). SOLOMON: a method for splitting a sample into equivalent subsamples in factor analysis. *Behavior Research Methods*, 54, 2665-2677.
- Mann, T. (1994). Informed Consent for Psychological Research: Do Subjects Comprehend Consent Forms and Understand Their Legal Rights? *Psychological Science*, 5(3), 140-143.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256. <https://doi.org/10.1177/026553229601300302>
- Mickan, P., & Motteram, J. (2008). An ethnographic study of classroom instruction in an IELTS preparation program. In: J. Osborne (Ed.) *IELTS Research Reports* (Vol. 8, pp. 17-43). IELTS Australia.
- Mickan, P., & Motteram, J. (2009). The preparation practices of IELTS candidates: Case studies. In J. Osborne (Ed.) *IELTS Research Reports* (Vol. 10, pp. 223-262). British Council.
- Morrow, K. (1986). The evaluation of tests of communicative performance. In M. Portal (Ed.). *Innovations in language testing* (pp. 1-13). NEFR/Nelson.
- Moss, P. A. (2016). Shifting the focus of validity for test use. *Assessment in Education: Principles, Policy & Practice*, 23(2), 236-251. <https://doi.org/10.1080/0969594X.2015.1072085>
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33-49. <https://doi.org/10.1177/0265532209347148>
- Nguyen, M. H. T. (2020). Beyond the washback of the English National Exam on learning: Subsequent impact on language learning strategies of first-year English majors. *Language Education & Assessment*, 3(2), 38-57. <https://doi.org/10.29140/lea.v3n2.356>
- 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.
- Oxford, R. L. (2013). Understanding language learner narratives. In: J. Arnold, & T. Murphey (Eds.) *Meaningful action: Earl Stevick's influence on language teaching* (pp. 95-110). Cambridge University Press.

- Pan, Y. (2009). Voices in the field: An interview with Jessica Wu. *SHIKEN: The Japan Association of Language Teaching, Testing & Assessment SIG Newsletter*, 13, 9-14.
- Pan, Y. C. (2015). Test impact: English certification exit requirements in Taiwan. *TEFLIN Journal*, 20(2), 119-139. <https://doi.org/10.15639/teflinjournal.v20i2/119-139>
- 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/10.22329/jtl.v9i1.3540>
- Pearson, W. S. (2019). Critical perspectives on the IELTS test. *ELT Journal*, 73(2), 197-206. <https://doi.org/10.1093/elt/ccz006>
- Phuong, H. Y. (2017). Improving English language teaching in Vietnam: Voices from university teachers and students. *Current Politics and Economics of South, Southeastern, and Central Asia*, 26(3), 285-310.
- Popham, W. J. (2014). The right test for the wrong reason. *Phi Delta Kappan*, 96(1), 46-52. <https://doi.org/10.1177/0031721714547862>
- Prodromou, L. (1995). The backwash effect: From testing to teaching. *ELT Journal*, 49(1), 13-25. <https://doi.org/10.1093/elt/49.1.13>
- Read, J., & Hayes, B. (2003). The Impact of IELTS on Preparation for Academic Study in New Zealand. In R. Tulloh (Ed.) *IELTS Research Reports* (Vol. 4, pp.153-206). IELTS Australia.
- 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 & Learning*, 6(3), 243-259. <https://doi.org/10.5172/ijpl.2011.6.3.243>
- Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23(3), 197-227. <https://doi.org/10.3102/01623737023003197>
- Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. *Theory and Research in Education*, 7(2), 224-233. <https://doi.org/10.1177/1477878509104327>
- Saville, N. D. (2009). *Developing a Model for Investigating the Impact of Language Assessment within Educational Contexts by a Public Examination Provider*. Unpublished doctoral dissertation, University of Bedfordshire.
- Shao, H. (2006). An empirical study of washback from CET-4 on college English teaching and learning. *CELEA Journal*, 29(1), 54-59.
- Shih, C. M. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), 135-161. <https://doi.org/10.3138/cmlr.64.1.135>
- Shih, P. C. (2013). *The English Benchmark Policy for Graduation: An Investigation of Perception, Motivation, and Approaches to Learning at a University of Technology in Central Taiwan*. Unpublished doctoral dissertation, Durham University.
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. NFLC occasional papers. <https://files.eric.ed.gov/fulltext/ED362040.pdf>.

- Shohamy, E. (2001). *The power of Tests: A Critical Perspective on the Uses of Language Tests*. Longman and Pearson Education Limited.
- Shohamy, E. (2006). *Language Policy: Hidden Agendas and New Approaches*. Routledge.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Stecher, B. (2010). *Performance Assessment in an Era of Standards-based Educational Accountability*. Stanford University, Stanford Center for Opportunity Policy in Education.
- Stoneman, B. (2006). *The impact of an exit English test on Hong Kong undergraduates: A study investigating the effects of test status on students' test preparation behaviours*. <https://theses.lib.polyu.edu.hk/handle/200/5489>.
- Su, M. H. M. (2005). A study of EFL technological and vocational college students' language learning strategies and their self-perceived English proficiency. *Electronic Journal of Foreign Language Teaching*, 2(1), 44-56.
- Sun, Y. (2016). *Context, construct, and consequences: Washback of the College English Test in China*. <https://qspace.library.queensu.ca/handle/1974/13985>.
- Thorpe, A., Snell, M., Davey-Evans, S., & Talman, R. (2017). Improving the academic performance of nonnative English-speaking students: the contribution of pre-sessional English language programmes. *Higher Education Quarterly*, 71(1), 5-32. <https://doi.org/10.1111/hequ.12109>
- Tsagari, D. (2007). *Review of washback in language testing: What has been done? What more needs doing?*. <https://files.eric.ed.gov/fulltext/ED497709.pdf>
- Tsai, Y., & Tsou, C. H. (2009). A standardised English language proficiency test as the graduation benchmark: Student perspectives on its application in higher education. *Assessment in Education: Principles, Policy & Practice*, 16(3), 319-330. <https://doi.org/10.1080/09695940903319711>
- Wall, D. (1997). Impact and washback in language testing. In C. Clapham, & D. Corson (Eds.). *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 291-302). Kluwer Academic.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled?. *System*, 28(4), 499-509. [https://doi.org/10.1016/S0346-251X\(00\)00035-X](https://doi.org/10.1016/S0346-251X(00)00035-X)
- Wang, Y. H., & Liao, H. C. (2012). Anxiety of university students in Taiwan about the General English Proficiency Test. *Social Behavior and Personality: An international journal*, 40(1), 63-74. <https://doi.org/10.2224/sbp.2012.40.1.63>
- Warden, C. A., & Lin, H. (2000). Existence of integrative motivation in an Asian EFL setting. *Foreign Language Annals*, 33(5), 535-545. <https://doi.org/10.1111/j.1944-9720.2000.tb01997.x>
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe & A. Curtis (Eds.) *Washback in language testing: Research contexts and methods* (pp. 19-36). Lawrence Erlbaum.

- Woodrow, L. (2011). College English writing affect: Self-efficacy and anxiety. *System*, 39(4), 510-522. <https://doi.org/10.1016/j.system.2011.10.017>
- Wu, J., & Lee, M. C. L. (2017). The relationships between test performance and students' perceptions of learning motivation, test value, and test anxiety in the context of the English benchmark requirement for graduation in Taiwan's universities. *Language Testing in Asia*, 7(9), 1-21. <https://doi.org/10.1186/s40468-017-0041-4>
- Xiao, W. (2014). The intensity and direction of CET washback on Chinese College students' test-taking strategy use. *Theory & Practice in Language Studies*, 4(6), 1171-1177. <https://doi.org/10.4304/tpls.4.6.1171-1177>
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10(2), 196-218. <https://doi.org/10.1080/15434303.2012.721423>
- 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). *A study on the washback effects of the Test for English Majors (TEM): Implications for testing and teaching reforms*. Springer.
- Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of class English learning: Insights from possible self theories. *Assessment in Education: Principles, Policy & Practice*, 21(1), 71-89. <https://doi.org/10.1080/0969594X.2012.757546>

Appendix A. Initial pool of items on the test factors

References at the bottom of the table.

Items	Justifications
1. I find the [TEST IN QUESTION] preparation materials difficult.	Saville (2009), Jones and Saville (2006), Moss (2016) underscored the communication of construct-related features in test preparation materials to give learners insights into test preparation practices and improve their learning.
2. Through comments made by students of the previous cohorts, I expect the [TEST IN QUESTION] to be difficult.	Schraw and Roedel (1994) found that the judgment of different test takers within a homogenous test accounted for the vast majority of variation in test difficulty.
3. Through comments made by lecturers, I expect the [TEST IN QUESTION] to be difficult.	Expert judgment had a role to play in predicting for test takers the difficulty of oral language proficiency tests (Elder et al., 2002), essay tests (HampLyons and Mathias, 1994) as well as reading and listening tests (Choi and Moon, 2019).
4. I think that the required passing score on the [TEST IN QUESTION] will be difficult to achieve.	Chen and Squires (2010), Wang and Liao (2012) found that the high required passing score may exert pressure on learners during test preparation.
5. I think that the [TEST IN QUESTION] is more difficult than the other test alternatives.	Difficulty of IELTS and TOEFL reading and listening subtests was compared in terms of cognitive skills involved (Baghaei et al., 2020), test content and method (Nguyen, 2018).
6. The time allotted to complete the [TEST IN QUESTION] is too short.	
7. I think that the language used in questions of the [TEST IN QUESTION] test paper is complex to understand.	Hawkey (2005) referred to (a) time pressure, (b) difficulty of language used, (c) difficulty of question, and (d) unfamiliarity with topics as the factors that caused difficulty to IELTS candidates.
8. I think that the questions in the [TEST IN QUESTION] test paper are difficult to answer.	
9. The topics in the [TEST IN QUESTION] test paper are unfamiliar for me to deal with.	
10. I find the task types in the [TEST IN QUESTION] test paper difficult to handle.	Difficulty in language tests were found in task types for listening (Brindley and Slatyer, 2002), reading (Zhang and Hope, 2020), speaking (Fulcher and Reiter, 2003), writing (Hamp-Lyons and Prochnow, 1991).
11. The [TEST IN QUESTION] result decides if I can graduate on time.	According to Le (2017), Nguyen (2013), Phuong (2017), English language proficiency tests in general and the IELTS test in particular is part of the graduation benchmark policy in Vietnam
12. The [TEST IN QUESTION] result affects my future job opportunities.	
13. The [TEST IN QUESTION] result affects my employability for desired jobs that involve using language skills.	Employers rely on the IELTS test to evaluate whether their prospective employees can function in the workplace (Pearson, 2019)
14. The [TEST IN QUESTION] result shows whether I am qualified to work in an area that fits my university major.	

15. The [TEST IN QUESTION] result shows whether I am qualified to work in a multi-national company.	
16. The [TEST IN QUESTION] result affects my future opportunities for higher education in other countries.	
17. The [TEST IN QUESTION] result affects my choice for higher education institutions abroad.	The original purpose of the IELTS test is for the language selection process to “measure the English language skills of candidates intending to study in academic or training contexts in [English-medium programmes]” (Ingram, 2004, p.18)
18. The [TEST IN QUESTION] result affects the chance that I will get scholarships to pursue higher education abroad.	
19. The [TEST IN QUESTION] result affects my social identity as an educated person.	Shohamy (2006) indicated that “[i]n countries where English is not the major national language, it is knowledge of the powerful global language, English, that often serves as a class marker enabling entrance to power groups – [the urban, the wealthy and the highly-educated]” (p.144)
20. The [TEST IN QUESTION] result affects how people judge the outcomes of my time at university.	
21. The [TEST IN QUESTION] result helps me to gain employment after graduation.	
22. The [TEST IN QUESTION] result helps me to enhance my competitiveness in future job seeking.	
23. The [TEST IN QUESTION] result helps me to enhance my competitiveness in the future workplace.	Green (2019) argued that the IELTS certificate is a sought-after proof of English language proficiency to enhance the status and employability of students after graduation.
24. The [TEST IN QUESTION] result helps me to be better prepared for multicultural working environment after graduation.	
25. The [TEST IN QUESTION] result helps me to evaluate my own English language proficiency.	Pearson (2009) suggested that the IELTS test provides non-native English test takers with “simplified, easy-to-understand, criterion-referenced, and time-bound evidence of that person’s English proficiency” (p.199).
26. The [TEST IN QUESTION] result helps me to be clearer about where I should go next in English learning.	IELTS helps the test takers clearly define their goals and study harder (IELTS, 2017).
27. The [TEST IN QUESTION] result helps me to gain equal access to learning opportunities.	Considering the inequalities between social classes (Bourdieu and Passeron, 1990), Green (2019) mentioned the disadvantaged group of learners who fail to be offered “the educational opportunities that knowledge of English opens up” (p. 209).
28. The [TEST IN QUESTION] result helps me to get quality learning opportunities at higher education institutions overseas.	IELTS acts as a “global gatekeeping institution” that either rejects or accepts people for their pursuit of academic study overseas. (Pearson, 2019, p. 197).
29. The [TEST IN QUESTION] result is a foundation to improve my English language abilities when I pursue higher education in English-speaking countries.	Green (2019) talked about the valued advantage of certificates of English language proficiency to promise an enhancement in English language abilities when international students choose the USA, the UK and Australia as the popular destinations for their study.

30. The [TEST IN QUESTION] result helps me to better prepare to function well for higher education overseas. Thorpe et al. (2017) suggest that IELTS is trusted as a satisfactory indicator of the test takers' abilities to function in an academic environment

References for Appendix A

- Baghaei, S., Bagheri, M. S., & Yamini, M. (2020). Analysis of IELTS and TOEFL reading and listening tests in terms of Revised Bloom's Taxonomy. *Cogent Education*, 7, 1-23. <https://doi.org/10.1080/2331186X.2020.1720939>
- Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, 19(4), 369-394. <https://doi.org/10.1191/0265532202lt236oa>
- Choi, I., & Moon, Y. (2019). Predicting the Difficulty of EFL Tests Based on Corpus Linguistic Features and Expert Judgment. *Language Assessment Quarterly*, 17(3), 1-25. <https://doi.org/10.1080/15434303.2019.1674315>
- Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: What does the test-taker have to offer? *Language Testing*, 19(4), 347-368. <https://doi.org/10.1191/0265532202lt235o>
- Fulcher, G., & Reiter, R. M. (2003). Task difficulty in speaking tests. *Language Testing*, 20(3), 321-344. <https://doi.org/10.1191/0265532203lt259oa>
- Hamp-Lyons, L., & Mathias, S. P. (1994). Examining expert judgments of task difficulty on essay tests. *Journal of Second Language Writing*, 3(1), 49-68. [https://doi.org/10.1016/1060-3743\(94\)90005-1](https://doi.org/10.1016/1060-3743(94)90005-1)
- Hamp-Lyons, L., & Prochnow, S. (1991). The Difficulties of Difficulty: Prompts in Writing Assessment. In: A. Sarinee (Ed.). *Current Developments in Language Testing* (pp. 58-76). Regional Language Centre Seminar on Language Testing and Language Programme Evaluation.
- Hawkey, R. (2005). An impact study of a high-stakes test (IELTS): lessons for test validation and linguistic diversity. In: M. Milanovic, & C. J. Weir (Eds.). *Multilingualism and Assessment: Achieving transparency, assuring quality, sustaining diversity* (pp. 215-228). Cambridge University Press.
- IELTS. (2017). *The Added Benefits Gained by preparing to Take the IELTS Exam*. [Accessed 20 July 2021]. Available from: <https://ieltscanadatest.com/2017/05/the-added-benefits-gained-by-preparing-to-take-the-ielts-exam/>.
- Nguyen, T. N. H. (2018). Test takers' attitudes to the test content of the two listening tests: IELTS AND TOEFL iBT. *VNU Journal of Foreign Studies*, 34(4), 36-58. <https://doi.org/10.25073/2525-2445/vnufs.4280>
- Phuong, H. Y. (2017). Improving English language teaching in Vietnam: Voices from university teachers and students. *Current Politics and Economics of South, Southeastern, and Central Asia*, 26(3), 285-310.
- Schraw, G., & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory and Cognition*, 22(1), 63-69.
- Zhang, R., & Hope, J. (2020). What is the Difficulty level of Question Types in the IELTS Academic Reading Sub-test?. *International Journal of Social Policy and Education*, 2(7), 45-50.

Appendix B. Initial pool of items on the IELTS washback on learning

Items	Justification
1. Because of the [TEST IN QUESTION], I see the reasons for learning English.	Nhan (2013) suggested that the high-stakes English Language Proficiency Tests (ELPT) give students positive perceptions of what they learn English for, i.e., future career and further education.
2. Because of the [TEST IN QUESTION], I set goals to develop English language skills.	
3. Because of the [TEST IN QUESTION], I have become more interested in learning English.	
4. Because of the [TEST IN QUESTION], I have more positive attitudes towards the English language.	Li (1990), Su (2005), Pan (2009), Pan and Newfield (2012) found the relationship between ELPTs and students' goal, interest, attitudes, efforts in improving their English language skills.
5. Because of the [TEST IN QUESTION], I make effort to improve English skills in English classes.	Tsai and Tsou (2009), Chu and Yeh (2017), Wu and Lee (2017) found that the high-stakes ELPTs motivated students' sustained actions driven from the provoked needs inside to improve their English competences.
6. Because of the [TEST IN QUESTION], I make effort to improve English skills outside English classes.	
7. Because of the [TEST IN QUESTION], I feel more personally committed to improving my English language competences.	
8. I set goals to achieve the targeted [TEST IN QUESTION] score.	Thanks to the implementation of high-stakes ELPT, students wanted to have the sense of accomplishment in their learning (Pan, 2015).
9. I have a desire to achieve the targeted [TEST IN QUESTION] score.	
10. I make effort to achieve the targeted [TEST IN QUESTION] score.	The high-stake ELPT motivated test takers to devote more intense efforts in pursuit of the desired test performance (Deci & Ryan, 2016; Cavendish et al., 2017).
11. Because of the IELTS, I feel pressure to learn English.	High-stakes ELPT engender feelings of negativity such as stress and worries (Chen, 2012, Woodrow, 2011; Damer and Malendres, 2011).
12. I am worried that my anxiety about the [TEST IN QUESTION] will affect my performance in the [TEST IN QUESTION].	
LA13. Because of the [TEST IN QUESTION], I feel discouraged from learning English	Roderick and Engel (2001), Pan and Newfields (2012) suggested that high-stakes ELPT made the test takers too anxious and consequently, get discouraged and abandon learning efforts.
LA14. I feel the [TEST IN QUESTION] is a burden to me.	Pan and Newsfield (2012) found that low-level English proficiency learners might find it burdensome to study for the ELPT.
LA15. I have a lot of self-doubt about my ability to pass the [TEST IN QUESTION]	Students experienced worries about failure in the test and ended up becoming pessimistic about their ability (Roderick & Engel, 2001).
LA16. My worries about the [TEST IN QUESTION] make me think passing the [TEST IN QUESTION] is unachievable.	Pan and Newsfield (2012) found that low-level English proficiency learners might find it impossible to pass the ELPT for graduation.

HL17. Because of the [TEST IN QUESTION], I spend lots of time developing my English language skills.	Students in preparation for the IELTS test engaged in practicing productive skills for the target language use domain (Allen, 2006).
HL18. Having to take the [TEST IN QUESTION] makes me aware of my weaknesses in my English language competence.	
HL19. Because of the [TEST IN QUESTION], I set clearer learning objectives.	High-stakes ELPT enable students to reflect on their learning progress by understanding the level they are at and perceiving areas of improvement in their knowledge and skills (Stecher, 2010; Popham, 2014).
HL20. Because of the [TEST IN QUESTION], I deal with gaps in my English language knowledge.	
HL21. Because of the [TEST IN QUESTION], I deal with gaps in my English language skills.	
HL22. Because of the [TEST IN QUESTION], I seek help from my tutor to guide my own learning.	Mickan and Motteram (2009) mentioned some students' dependency on expert help to develop personal agency and strategic action in preparation for the IELTS test.
RL23. My learning activities are only for passing the [TEST IN QUESTION].	Chapelle et al. (2011), Brown and Abeywickrama (2010) talked about the precedence taken over achieving the desirable test score rather than learning to develop English language skills.
RL24. My learning is narrowed to practicing mock tests to achieve the targeted [TEST IN QUESTION] score.	Ren (2011) investigating CET-4 in China, Mickan and Motteram (2008) investigating the IELTS found that student's learning was narrowed with the focus on mock tests.
RL25. I consult my tutor only on test-taking tips to achieve the targeted [TEST IN QUESTION] score.	Mickan and Motteram's (2009), Xie and Andrews (2013) mentioned social strategies adopted by some learners through asking a knowledgeable mentor for some test-taking tips on how to achieve the targeted score
RL26. I memorize sample answers to the questions that are likely to appear in the real [TEST IN QUESTION].	Ren (2011) found that the students in preparation for the CET-4 test memorized samples of writing prompts.
RL27. My learning is narrowed to learning test-taking tips to achieve the targeted [TEST IN QUESTION] score.	Xiao (2014) also found the negative test-wiseness strategy adopted by the students in getting their answers to test items by exclusively relying on test characteristics.

Appendix C. Factor structure of the test factors instrument

Items of the test factors	Factor loadings		
<i>Factor 1. Test difficulty ($\alpha = 0.70$)</i>			
TD4. I think that the required passing score on the IELTS will be difficult to achieve.	0.776	-0.032	-0.311
TD5. I think that the IELTS is more difficult than the other test alternatives.	0.665	0.014	-0.096
TD6. The time allotted to complete the IELTS is too short.	0.582	-0.145	0.348
TD8. I think that the questions in the IELTS test paper are difficult to answer.	0.509	0.213	0.227
<i>Factor 2. Test stakes ($\alpha = 0.72$)</i>			
TS11. The IELTS result decides if I can graduate on time	0.027	0.804	0.029
TS13. The IELTS result affects my employability for desired job vacancies.	-0.031	0.750	0.209
TS14. The IELTS result shows whether I am qualified to work in an area that fits my university major.	0.061	0.661	-0.201
TS17. The IELTS result affects my choice for higher education institutions abroad.	-0.155	0.638	-0.144
TS20. The IELTS result affects how people judge the outcomes of my time at university.	0.034	0.488	0.337
<i>Factor 3. Test utility ($\alpha = 0.74$)</i>			
TS21. The IELTS result helps me to gain employment after graduation.	0.013	-0.094	0.879
TS22. The IELTS result helps me to enhance my competitiveness in future job seeking.	0.053	0.135	0.790
TS26. The IELTS result helps me to be clearer about where I should go next in English language learning.	0.043	-0.125	0.754
TS27. The IELTS result helps me to gain equal access to learning opportunities.	-0.222	0.006	0.724

Appendix D. Factor structure of the IELTS washback on learning

Items of the IELTS washback on learning	Factor loadings			
Factor 1. Learning motivation ($\alpha = 0.74$)				
LM2. Because of the IELTS, I set goals to develop English language skills.	0.865	-0.037	0.008	0.004
LM6. Because of the IELTS, I make effort to improve English skills outside English classes.	0.698	0.102	0.205	-0.012
LM8. I set goals to achieve the targeted test score.	0.619	0.280	-0.097	0.346
Factor 2. Learning anxiety ($\alpha = 0.70$)				
LA11. Because of the IELTS, I am under pressure to learn English.	0.134	0.860	-0.328	0.105
LA15. Because of the IELTS, I feel discouraged from learning English.	-0.107	0.717	0.013	-0.038
LA18. I have a lot of self-doubt about my ability to pass the IELTS.	-0.322	0.487	0.377	0.100
Factor 3. Holistic learning ($\alpha = 0.76$)				
HL21. Because of the IELTS, I spend lots of time developing my English language skills.	-0.008	-0.016	0.876	0.170
HL22. Having to take the IELTS makes me deal with weaknesses in my English language competence.	-0.083	0.024	0.573	0.078
HL27. Because of the IELTS, I seek help from my tutor to guide my own learning.	0.320	0.052	0.529	0.004
Factor 4. Restricted learning ($\alpha = 0.76$)				
RL29. I focus on practicing mock tests to achieve the targeted IELTS score.	0.144	-0.105	-0.035	0.833
RL31. I memorize sample answers to the questions that are likely to appear in the real test.	0.195	0.027	-0.028	0.762
RL32. I focus on learning test-taking tips to achieve the targeted IELTS score.	-0.272	0.311	-0.072	0.729
RL33. I focus on practicing test-related material to achieve the targeted IELTS score.	0.365	0.010	0.025	0.534