LEARN Journal: Language Education and Acquisition Research Network
ISSN: 2630-0672 (Print) | ISSN: 2672-9431 (Online)
Volume: 18, No: 1, January – June 2025

Language Institute, Thammasat University
https://so04.tci-thaijo.org/index.php/LEARN/index

Washback of Vietnamese Standardized Test of English Proficiency (VSTEP.3-5) on Undergraduate Students' Learning Strategies

Phuong Ai Ngoc Doan[a], Chatraporn Piamsai[b,*]

[a]dnaphuong@hueuni.edu.vn, English as an International Language, Graduate School, Chulalongkorn University, Thailand
[b] chatraporn.p@chula.ac.th, Chulalongkorn University Language Institute, Chulalongkorn University, Thailand
[*] Corresponding author, chatraporn.p@chula.ac.th

APA Citation:
Doan, N. A. P., & Piamsai, C. (2025). Washback of Vietnamese Standardized Test of English Proficiency (VSTEP.3-5) on undergraduate students' learning strategies. *LEARN Journal: Language Education and Acquisition Research Network, 18* (1), 23-53. https://doi.org/10.70730/VEZK2760

Received
06/05/2024

Received in revised form
19/07/2024

Accepted
22/07/2024

ABSTRACT

Studies on washback of not only international standardized proficiency tests, but also localized high-stakes tests have been increasingly attracting considerable interest. The current study explores the washback effects of a Vietnam-made localized English proficiency test named "Vietnamese Standardized Test of English Proficiency, level 3-5" (VSTEP.3-5) on undergraduate students' learning strategies. This research employed a mixed-methods design featuring two major research instruments labeled as questionnaires and semi-structured interviews. Quantitative data were collected from the survey questionnaires delivered to 600 English-majored sophomores at a public university in Central Vietnam. Additionally, qualitative data were obtained and analyzed from semi-structured interviews with 12 student participants through a prudent coding process. The findings indicated that a majority of undergraduate students claimed that VSTEP.3-5 affected their learning strategies in a positive way, especially

"cognitive and metacognitive strategies". Besides, preferences for distinct strategies utilized in specific English skills were particularly evident in such productive skills as Speaking and Writing.

Keywords: washback, localized high-stakes test, VSTEP.3-5, learning strategies, undergraduate students

Introduction

Washback – one of the six crucial attributes of test usefulness, "seen as the link between testing, teaching and learning" (Hamp-Lyons, 1997, Shohamy et al., 1996; as cited in Tsagari, 2009, p.4), has garnered growing attention from language testing experts, educators, and academics alike. According to Bachman and Palmer (1996, p.30), washback is regarded as one dimension of test impact, "the direct impact of testing on individuals" such as students and teachers, which can be either positive (beneficial) or negative (harmful). This can be further elaborated by distinguishing between "impact" and "washback". Whilst "impact" refers to the broader scale, involving the influence of a test on the society, educational system, as well as stakeholders, "washback" pertains to the narrower scope, referring to the consequences of a test on the process of teaching and learning. Upon closer examination, research into "washback" thoroughly investigates the influence of a test on teachers' instructional practices and on students' learning factors. In so doing, washback studies are not only significant, but also essential in a sense that any "negative washback" of a test on teaching and learning has the potential to be promptly detected and diminished.

Regarding the relationship between a language test and students' characteristics, language learning strategies have a vital role to play as learners "have greater potential to achieve success in language tests" as long as they "are equipped with necessary strategies" (Sadeghi et al., 2021, p.3). However, while washback research on such learners' traits as motivation and anxiety are numerous, e.g., motivation and autonomy (Buyukkeles, 2016; Dawadi, 2020; Lertcharoenwanich, 2020; Nguyen, T. L., 2017; Sadeghi, 2021), and anxiety (Brown, 1997; Dawadi, 2020; Shohamy, 2016), a modest number of publications have been found in regard to washback on learning strategies. More often than not, research into this area treated this factor as a minor part or fused them into the so-called "test-taking strategies" (Aftab et al., 2014; Ahmed, 2018; Green, 2007; Jamalifar et al., 2021; Sukyadi & Mardiani, 2011). On the one hand, Green (2007) implied that the intense link between a test and learning behavior is that test design can lure language learners to enhance test-taking strategies instead of developing lifelong learning competence. It

was also revealed in Sukyadi and Mardiani's (2011) study that the test negatively affected students' learning strategies in a way that students paid their most attention to classroom content relevant to the exam as well as test-taking strategies. As a result, "cramming tends to be adopted by most students in order to perform well in tests, rather than engaging in a real learning process" (Ahmed, 2018, p. 214). On the other hand, the study on the washback effects carried out by Fan and Yang (2020) revealed more positive outcomes which referred to "memory strategies" as the most commonly used strategy, closely followed by "metacognitive strategies." Conversely, the most recent study by Chak (2023) indicated that "cognitive and metacognitive strategies" were more commonly utilized by the students than "memory strategies." It is essential that more empirical studies on this trendy topic be in need of implementation, especially in EFL settings.

Furthermore, concerning the context of the study and the localized high-stakes test itself, a brief overview of VSTEP.3-5 is noteworthy. In Vietnam, thanks to tremendous amount of effort in 'educational reforms' (Doi Moi), the role of English as a foreign language in education has been firmly asserted thus far. With the ultimate goal for achieving "comprehensive reform of teaching, learning, and assessment of foreign languages at all education levels", the National Foreign Language 2020 Project has made one primary accomplishment, which is the provision of CEFR-VN, "a six-level foreign language competency framework aligned to the Common European Framework of Reference for Languages" in 2012. This led to the introduction of VSTEP.3-5, abbreviated for "Vietnamese Standardized Test of English Proficiency, level 3-5," or "Vietnamese three-level test of English proficiency targeted at Levels 3 to 5 of the CEFR-VN in early 2015" (Dunlea et al., 2018; MOET, 2015b; as cited in Vo, 2021, p.3). The VSTEP.3-5 test is aimed at assessing English proficiency from level 3 to level 5 according to the Common European Framework of Reference for languages for Vietnamese learners (CEFR-VN) or from level B1 to level C1 according to the Common European Framework of Reference for Languages (CEFR) for users in various majors and professions using four skills, emphasizing academic ones. In essence, VSTEP.3-5 is a high-stakes test that meets domestic demands, reflecting 'localization,' like China's CET, Taiwan's GEPT, or Japan's EIKEN. It is believed to be localized to "suit Vietnamese culture and language use." This localization is "built up through the process of contextual mediation in test design and development" (Saville, 2009, 2010; Wu, 2016; as cited in Nguyen, 2020, p. 82), including such factors as test content, task types, and test specifications. By virtue of the Ministry of Education and Training's policy, the position of VSTEP tests has been gradually asserted as "national English assessment instruments" and "alternative to the existing

expensive international standardized English tests (e.g. IELTS, TOEFL)" (Nguyen, 2018, p. 115).

Nonetheless, studies into washback effects of VSTEP.3-5 on learning strategies have yet to receive due attention, in spite of its significance in test development and validation. In other words, few of those studies examined the learners' use of learning strategies in close relation with washback effects. For instance, one of the recent studies on the subject in Vietnam, the study by Nguyen (2017), with the highest number of citations, solely put an emphasis on students' motivation and anxiety. Likewise, Bui (2018) shared the same concerns in her study on the washback effects of VSTEP level 3 on students' attitudes, motivation, and test anxiety. All things considered, research into washback effects of VSTEP.3-5 on students' use of learning strategies has remained a not widely explored area. Hence, this present study was implemented to explore a more insightful glimpse into the washback effects of VSTEP.3-5 on Vietnamese undergraduate students' use of language learning strategies.

Literature Review

Concepts of Washback

There have been a tremendous number of researchers who made attempts in defining "washback" from their own perspectives thus far. Bachman and Palmer (1996) regarded "washback" as one dimension of test impact, reflecting the micro level such as the effect of a test on individual students and teachers, which can be either positive or negative. "Impact," on the other hand, are testing consequences that a test has "on individuals (teachers and students), educational systems, and the society at large." (Bachman & Palmer, 1996, p. 30). Sharing the same viewpoint, both Wall (1997) and Hamp-Lyons (1997) distinguished between test impact and test washback in terms of the scope of the effects. Wall (1997, p. 291) claimed that whereas the test impact indicates "any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole," washback focuses more attention to "the effects of tests on teaching and learning." Likewise, Hamp-Lyons (1997) emphasized that the difference between impact and washback lies in the scope of the effects of testing - "which gives us a view of test consequences falling between the narrower one of washback and the all-encompassing one of impact." (Hamp-Lyons, 1997; as cited in Rahimi et al., 2016, p. 7). McNamara (2000) reconfirmed these definitions when using "impact" and "washback" to differentiate between two levels of this phenomenon, with the former indicating the effects of tests on macro-levels

of education and society, while the latter regarding the effects of language tests on micro-levels of language teaching and learning inside the classroom.

Washback Models and Frameworks in Relation to Teaching and Learning

Known as 'fathers' of the washback framework, Alderson and Wall (1993) pioneered the development of 15 hypotheses to indicate that "there is no one-to-one relationship between tests, good or bad, and their effect on teaching and learning." (Gu, 2007, p. 22) (Table 1)

Table 1

Alderson and Wall's (1993) Washback Framework

H.1	A test will influence teaching.
H.2	A test will influence learning.
H.3	A test will influence **what** teachers teach.
H.4	A test will influence **how** teachers teach.
H.5	A test will influence **what** learners learn.
H.6	A test will influence **how** learners learn.
H.7	A test will influence the **rate** and **sequence** of teaching.
H.8	A test will influence the **rate** and **sequence** of learning.
H.9	A test will influence the **degree** and **depth** of teaching.
H.10	A test will influence the **degree** and **depth** of learning.
H.11	A test will influence attitudes to the content, method, etc. of teaching and learning.
H.12	Tests that have important consequences will have washback.
H.13	Tests that do not have important consequences will have no washback.
H.14	Tests will have washback on **all** learners and teachers.
H.15	Tests will have washback effects for **some** learners and **some** teachers, but **not** for others.

(Alderson & Wall, 1993, p.120, original text in bold as indicated)

In the meantime, Hughes (1993) proposed a basic, yet influential process model of washback, "a model of washback mechanism" (Hughes, 1993; as cited in Gu, 2007, p. 24), which differentiated between effects on participants (the people affected by the test, e.g. teachers, learners, and materials writers), processes (participant actions, e.g. teaching and learning activities) and products (the outcomes of these processes: scores on tests, courses, teaching materials, etc.) (Xu & Liu, 2018).

A combination of Alderson and Wall's Washback Hypotheses and Huges' Washback Trichotomy was Bailey's (1996) basic model of washback. In her model, Bailey (1996) made a distinction between "washback to the program," which involves not only the teachers but all other participants except for the students, and "washback to the learners" (Bailey, 1996, pp. 264-265), emphasizing the influence of "test-derived information provided to the test-takers and having a direct impact on them" (Bailey, 1996, p. 263). Whilst the former is believed to contain five of Alderson and Wall's Washback Hypotheses (2, 5, 6, 8, and 10), the latter directly links to Hypotheses 1, 3, 4, 7 (Xu & Liu, 2018).

Unlike the aforementioned researchers who established the washback models chiefly around *participants*, Green (2007) took a different path by focusing on the dimensions of washback, namely *direction*, *intensity*, and *variability*. In his view, whether a test has negative or positive effects depends on its design and quality. *Variability* refers to the amount at which the washback varies on different participants, based on their "knowledge and understanding of participant values, motivation and resources" (Green, 2007; as cited in Green, 2013, p. 44). Meanwhile, the *intensity* of washback is determined by the "perceived importance and difficulty of the test" (Green, 2007, p. 14).

Key Concepts of Learning Strategies

In order to become a successful language learner, one may have their own strategy or way of learning to achieve the optimal outcomes. More and more attention has been drawn to research into language learning strategies than ever before, especially in the era of student-centeredness. However, it is far from straightforward for researchers to express unanimous approval of definitions and classifications of the term "learning strategies."

Brown (1980) found his own way to define such term by differentiating "learning strategies" from "communication strategies," stating that "communication is the output modality and learning is the input modality" (p. 87), yet it was he himself in subsequent years who admitted that "in the arena of linguistic interaction, it is sometimes difficult to distinguish between the two" (1994, p. 118). Besides, Ellis's (1994) five major aspects of good or successful learners are also worth noting, which are: "a concern for language form; a concern for communication; an active task approach; an awareness of the learning process; and a capacity to use strategies flexibly in accordance with task requirements." (p. 546).

Conceptual Framework for Learning Strategies

Oxford's (1990) Taxonomy

The most ubiquitous conceptual framework for learning strategies is the Strategy Inventory for Language Learning (SILL) developed by Oxford (1990). Operated in two versions, one for English speakers learning a new language (version 5.0) and the other for speakers of other language learning English (version 7.0), the SILL constitutes direct and indirect learning strategies depending on the extent to which each strategy item is involved in language learning. Specifically, compared to O'Malley and Chamot's (1990) taxonomy, Oxford's (1990) taxonomy provided a more detailed and comprehensive classification of language learning strategies, involving six following distinct categories: *Memory Strategies* (how students remember language), *Cognitive Strategies* (how students think about their learning), *Compensation Strategies* (enabling students to make up for limited knowledge), *Metacognitive Strategies* (how students manage their own learning), *Affective Strategies* (students' feelings) and *Social Strategies* (involving learning by interaction with others).

Cohen and Chi's (2004) Taxonomy

Another widely used language learning strategy inventory is the Learning Strategy Use Inventory (LSUI) designed by Cohen and Chi (2004). Unlike the SILL by Oxford (1990), the inventory is developed through classifying language learning strategies according to language skills and elements, covering six main areas namely "listening, vocabulary, speaking, reading, writing, and translation sections" as follows (Table 2).

Table 2

Cohen and Chi's (2004) Taxonomy

Listening Strategies	Vocabulary Learning Strategies	Speaking Strategies	Reading Strategies	Writing Strategies	Translation Strategies

• Strategies to increase my exposure to the TL. • Strategies to become more familiar with the sounds in the TL. • Strategies to prepare to listen to conversation. • Strategies to listen to conversation in the TL.	• Strategies for when I do not understand some or most of what someone says in the TL. • Strategies to learn new words. • Strategies to review vocabulary. • Strategies to recall vocabulary. • Strategies to make use of new vocabulary.	• Strategies to practice speaking. • Strategies to engage in conversations. • Strategies for when I can't think of a word or an expression.	• Strategies to improve my reading ability. • Strategies for when words and grammatical structures are not understood.	• Strategies for basic writing. • Strategies for writing an essay or academic paper. • Strategies to use after writing a draft of an essay or paper	• Strategies for translation. • Strategies for working directly in the TL

(Cohen & Chi, 2004, p.19)

VSTEP.3-5

Target population

VSTEP.3-5, a multi-level test, is considered "the first variant of the VSTEP (Vietnamese Standardized Test of English Proficiency), aiming at levels 3 to 5 according to the CEFR-VN (or levels B1 to C1 according to the CEFR)". It is targeted at test-takers who are adults aged 18 and above, especially "English teachers, tertiary students, senior officials, and lecturers" (Vietnam's Ministry of Education & Training, 2015, p. 2). VSTEP.3-5 test emphasizes academic abilities rather than occupational ones. It is held around 4 or 5 times a year, with an agreeable test fee, compared to that of other international standardized English tests.

Test format

In terms of test format, VSTEP.3-5 is a pen-and-paper test assessing all four skills: listening (40 minutes), reading (60 minutes), writing (60 minutes), and speaking (12 minutes), in which the Reading section (40 questions) and the Listening section (35 questions) are in the forms of multiple choice selected items, while the Speaking section is a "face-to-face interview" with three parts and the Writing section is composed of two writing tasks (Task 1: email-writing; Task 2: essay-writing) (Appendix).

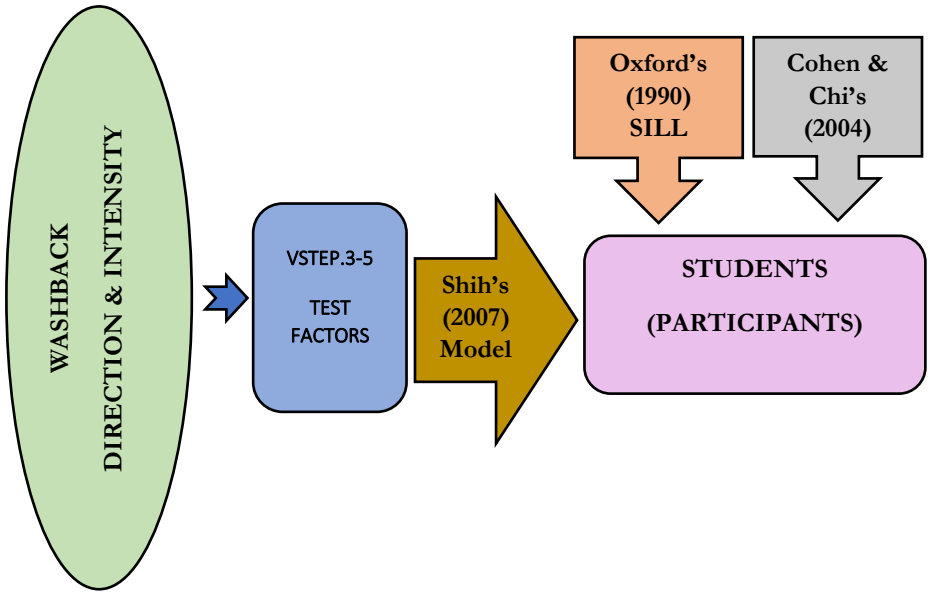
The Present Study's Conceptual Framework for Washback Effects of VSTEP.3-5 on Learning Strategies

The skeleton of the present study's washback model is chiefly built from Shih's (2007) washback model, in which such "test factors" as "(1) stakes of the test, (2) importance of the test, (3) content of the test, (4) difficulty of the test, (5) structure of the test, (6) format of the test, (7) tested skills, (8) purpose of the test" have washback effects on "participants" (students) and "processes" (students' learning strategies). To be more precise, Shih's (2007) model reassured the complexity of washback on learning based on Alderson and Wall's 15 Hypotheses (1993) and Bailey's (1996) Model. In his model, there was an inter-linkage among the three factors named "extrinsic factors, intrinsic factors, test factors," and they all had their impact on "washback of a test on student's learning and psychology." (Shih, 2007, p. 151). Also, Shih pointed out five areas of learning that can be affected by a test, involving (1) content of learning, (2) total time of learning, (3) learning strategies, (4) learning motivation, and (5) test anxiety. Since Shih's (2007) model only refers to "learning strategies" as one of the five students' factors affected by the test, there is an urgent need for a more in-depth model featuring the specific types of learning strategies that a test may affect. Besides, Oxford's (1990) taxonomy SILL was employed in the current study as the conceptual framework to explore the washback effects of VSTEP.3-5 on strategy use. The reasons are twofold: first, the SILL has been a commonly used instrument across a large and various groups of language and cultural backgrounds (Oxford, 1996); second, its reliability and validity has been ensured (Oxford, 1992; as cited in Khalil, 2005, p. 110). Most importantly, Oxford's taxonomy with SILL has been widely adopted in many studies related to washback effects. In the present study, Oxford's taxonomy was adopted in combination with Cohen and Chi (2004)'s taxonomy LSUI to be compatible with VSTEP.3-5 test constructs and specific language skills (reading skill, listening skill, speaking skill, and writing skill). Last but not least, Green's (2007) two major dimensions of washback namely "direction"

(whether washback is positive or negative) and "intensity" (whether washback is intense or weak) were also adopted in the present study's framework (Figure 1).

Figure 1

The Present Study's Conceptual Framework



Methodology

Research Design

The present study employed a mixed-methods explanatory research design to elicit both quantitative and qualitative data, yet more attention was drawn onto the qualitative approach. This research design accords with Hawkey when he claimed that the nature of washback studies is "complex, non-monolithic and iterative," and "likely to show influences from both ends of Lazaration's continua, but perhaps with an inclination towards the qualitative ends" (Hawkey, 2006, p. 30).

Research Site

The study was carried out at a recognized public university in Central Vietnam. The reason for choosing the said research site lies in its prestige and

representation. First of all, it is among the top ten selected universities and institutions authorized by Vietnam's MOET to issue VSTEP certificates (Vietnam's Decision 23/2017/TT-BGD-DT, 2017). Students at this university can represent Vietnamese EFL students, or at least those in the Central Region of Vietnam. Regarding the university policy, obtaining a C1 level proficiency certificate is one of the English Language Program's graduation prerequisites. English-majored students at the said university are required to take the VSTEP.3-5 exam and are expected to get the passing score of 8.5 out of 10 to reach level 5 of CEFR-VN, equivalent to C1 level of CEFR.

Participants

The participants were the second year English-majored undergraduate students in the first semester, academic year 2022 at a university in Central Vietnam. The number of respondents to questionnaires and the number of interviewees was 600 (Female = 539, Male =49, Others = 12) and 12 (Female = 10, Male = 2), respectively.

General English Skills Courses (Listening, Speaking, Reading, and Writing) are considered their compulsory courses offered by the Faculty of English. In terms of the student participants' levels of English proficiency, all student participants were sophomores with at least 7 years of learning English and had passed the English Final Exams after one year studying at Faculty of English. The test used for assessing their English Proficiency at the end of the semester is the standardized test at the B2 level of English in the Common European Framework of Reference (CEFR), equivalent to level 4 according to CEFR-VN.

Research Instruments

Questionnaires

Questionnaire items, consisting of 52 items, were generated based on the operational definitions of the aspects to be measured. Some items were borrowed from existing Language Learning Strategies questionnaires (Chen, 2002; Oxford, 1990; Hawkey, 2006; Nguyen & Gu, 2020; Nguyen, 2020; Wall, 2005) and modified to make them suitable for the current study. The questionnaire items were written in both English and Vietnamese.

Semi-structured interviews

In this study, this type of interviewing technique was conducted at the end of the semester. Interview questions were divided into 4 clusters, namely: demographic questions, experience/ behavior questions, opinion/value questions, and feeling questions. Some questions were constructed and modified from Hawkey (2006), Shih's (2007) model and Pan's (2013) model. The language use of semi-structured interviews was merely Vietnamese, the participants' mother tongue, to avoid confusion and misunderstanding.

Demographic information about selected participants is presented below (Table 3).

Table 3

Characteristics of Participants for Interviews

Participants' ID.	Gender	Age	Years of learning English	Level of English proficiency
S01	Male	19	10 years	Lower-intermediate
S02	Female	20	10 years	Intermediate
S03	Female	19	8-10 years	Intermediate
S04	Female	19	>10 years	Intermediate
S05	Female	19	>10 years	Upper-Intermediate
S06	Female	19	>10 years	Intermediate
S07	Female	19	>10 years	Intermediate
S08	Female	20	10 years	Intermediate
S09	Female	19	>10 years	Intermediate
S10	Female	19	8-10 years	Lower-intermediate
S11	Male	19	10 years	Intermediate
S12	Female	19	10 years	Intermediate

Reliability of research instruments

In order to ensure the reliability of the questionnaires, data collected from the pilot study were used to measure its reliability coefficient, using SPSS. According to UCLA Institute for Digital Research & Education Statistical Consulting, a reliability coefficient of .70 or higher is considered "acceptable" in most social science research situations. Based on the data gathered for the piloted questionnaires, the Cronbach's Alphas is 0.912, which is highly acceptable.

Validation of research instruments

The research instruments were validated through a rigorous process, involving the participation and consultation of a group of five experts in the field from the U.S, Thailand, and Vietnam. They are experienced university lecturers whose expertise is Language Evaluation and Assessment. Concerning the validation tool, an index of item-objective congruence (IOC) was used for validating the constructs and the content of the research instruments.

Data Collection

Questionnaires were delivered to 600 sophomores in the middle of the semester. In reality, all 600 student questionnaires were fully answered by the student participants. Regarding the data collection via interviewing, 12 students were chosen purposefully for in-depth interviews at the end of the semester, based on their level of proficiency, the extent of their knowledge about the test, and their willingness to share opinions. Each face-to-face interview, which lasted for approximately 15 minutes, was recorded carefully.

Data Analysis

The data collected from questionnaires were analyzed through descriptive statistics. In so doing, the arithmetic mean and the standards deviation were determined with the help of Excel and SPSS Software (short for "Statistical Package for the Social Sciences") version 28.0. The interpretation of mean scores was based on the aforementioned Likert-scales in Research Instruments Section. Different scales led to different ways of interpretation, yet overall, there is a common range of mean scores as follows:

- 4.21 - 5.00 means "always"/ "always or almost always true of me"/ "increased greatly"/ "strongly agree"
- 3.41 - 4.20 means "often"/ "usually true of me"/ "increased slightly"/ "agree"
- 2.61 - 3.40 means "sometimes"/ "somewhat true of me"/ "remained the same"/ "neither agree nor disagree"
- 1.81 - 2.60 means "seldom"/ "usually not true of me"/ "decreased slightly"/ "disagree"
- 1.00 - 1.80 means "never"/ "never or almost never true of me"/ "decreased greatly"/ "strongly disagree"

The qualitative data obtained from interview recordings were initially transcribed verbatim by a Vietnamese AI-based application called "Memobot" – a useful tool to transcribe voices and convert audios into texts, then content analysis based on coding transcriptions was exploited to analyze the data thanks to NVivo Software version 14.0, a qualitative data analysis (QDA) computer software package produced by QSR International. As soon as essential codes, themes, example quotes were figured out, they were translated into English under the supervision of a professional translator. In short, by utilizing the aforementioned audio-to-text conversion tool, combined with the coding software and third-party translation support, the credibility and validity of the coding process could be guaranteed.

Issues of ethical consideration

The current study was carried out with the consent of all participants. A brief introduction to the research study and research objectives was given at the beginning of questionnaires and interview questions in the form of "Information Sheet for Research Participants." Every respondent to interviews had the right to read carefully and sign a "Letter of Consent" right before each interview. In order to ensure the participants' confidentiality, the real names of the participants were removed and coded in the numbers. Finally, all research instruments were rigorously reviewed and approved by "The Research Ethics Review Committee for Research Involving Human Subjects," using international standards.

Results

The quantitative data obtained from the questionnaires and the qualitative data from the interview transcriptions are presented sequentially, with the former being expressed mainly through numbers and tables, while the latter being illustrated through featured themes and interview excerpts.

Findings from quantitative data

Students' Previous Experiences Relevant to Language Learning Strategies Before the Introduction of VSTEP.3-5

Table 4

Students' Previous Experiences Relevant to Language Learning Strategies Before the Introduction of VSTEP.3-5

Item Statistics				**Interpretation**
Before the introduction of VSTEP.3-5	**Mean**	**Std. Deviation**	**N**	
I'd listen to varieties of English accents.	2.7033	1.03623	600	Sometimes
I'd make plans before doing an English task.	3.0900	1.02648	600	Sometimes
I'd monitor and evaluate an English task by myself.	2.9883	1.00493	600	Sometimes
I'd develop ideas from a topic in speaking.	3.5067	.82528	600	Often
I'd develop ideas from a topic sentence in Writing.	3.6400	.81127	600	Often
I'd focus on writing argumentative essays.	3.1083	.83931	600	Sometimes
I'd focus on writing emails or letters in English.	2.6500	1.12529	600	Sometimes
I'd make a point in English.	2.9183	1.08795	600	Sometimes
I'd provide supporting ideas and examples when writing.	3.8683	.77373	600	Often
I'd use other words to express ideas when I don't know/forget a word.	4.1833	.77074	600	Often
I'd paraphrase words/phrases/sentences when writing.	4.1883	.79829	600	Often
I'd guess the meaning of the words in context.	4.1167	.74430	600	Often
I'd make inferences from English reading texts.	3.5217	.83107	600	Often
I'd listen to announcements or radio broadcasts in English.	3.0083	.79539	600	Sometimes
I'd express my co-operation when working in pairs/groups.	3.8933	.77003	600	Often

(*) Interpretation was based on *scale of frequency: Never (1), Seldom (2), Sometimes (3), Often (4), and Always (5)*

It can be clearly seen from Table 4 that based on the students' reflections before the introduction of VSTEP.3-5 test, the most frequently used learning strategies belonged to "compensation strategies" according to Oxford's taxonomy (1990), including "paraphrasing words/phrases/sentences when writing" (Mean = 4.18), "using other words to express ideas when not knowing/forgetting a word" (Mean =4.18), followed by "guessing the meaning of the words in context" (Mean = 4.11). "Social strategies" (Item: "expressing co-operation when working in

pairs/groups"), and "cognitive strategies" (Item: "providing supporting ideas and examples when writing"), ranked second and third place, with Mean equivalent to 3.89 and 3.86, respectively. Though less often used, these following "cognitive strategies" still had moderate Mean values, namely "developing ideas from a topic sentence in Writing" (Mean = 3.64), "making inferences from English reading texts" (Mean = 3.52), "developing ideas from a topic in Speaking" (Mean = 3.50). Additionally, five "metacognitive strategies" that were "sometimes" used involves "writing argumentative essays" (Mean = 3.01), "making plans before doing an English task" (Mean = 3.09), "listening to announcements or radio broadcasts in English" (Mean = 3.00), "monitoring and evaluating an English task by myself" (Mean = 2.98), and "making a point in English" (Mean = 2.98). Last of all, the two least used "cognitive strategies" were "listening to varieties of English accents" (Mean = 2.70) and "writing emails or letters in English" (Mean = 2.65).

Perceived impact on students' learning strategies in and outside classroom activities

Table 5

Perceived Impact on Students' Learning Strategies in and outside Classroom Activities

Item Statistics				
After the introduction of VSTEP.3-5	**Mean**	**Std. Deviation**	**N**	**Interpretation**
I practice listening to varieties of English accents.	4.2117	.81320	600	Usually true of me
I make plans before doing an English task.	4.1433	.74187	600	Usually true of me
I monitor and evaluate an English task by myself.	3.7167	.71688	600	Usually true of me
I learn how to develop ideas from a topic in *speaking.*	4.3133	.70189	600	Always or almost always true of me
I learn how to develop ideas from a topic sentence in *writing.*	4.4483	.60658	600	Always or almost always true of me
I focus on writing argumentative essays.	4.1667	.76148	600	Usually true of me
I focus on writing emails or letters in English.	4.3067	.77263	600	Always or almost always true of me
I practice making a point in English.	3.9600	.70656	600	Usually true of me
I practice providing supporting ideas and examples when writing.	4.0367	.69001	600	Usually true of me

I use other words to express ideas when I don't know/forget a word.	3.9033	.64396	600	Usually true of me
I practice paraphrasing words/phrases/sentences when writing.	4.2333	.65544	600	Usually true of me
I practice guessing the meaning of the words in context.	4.2767	.65892	600	Usually true of me
I pay more attention to make inferences from English reading texts.	3.8783	.72185	600	Usually true of me
I take more notice of listening to announcements or radio broadcasts in English.	4.2117	.84738	600	Usually true of me
I express my co-operation when working in pairs/groups.	3.9050	.82083	600	Usually true of me

(*) Interpretation was based on *scale of reflection/likelihood: Never or almost never true of me (1), Usually not true of me (2), Somewhat true of me (3), Usually true of me (4), and Always or almost always true of me (5)*

As can be obviously seen in Table 5, there are remarkable increases in the extent of students' use of learning strategies after the introduction of VSTEP.3-5 test. Significant upward trends related to "cognitive strategies" were "developing ideas from a topic sentence in Writing" (Mean = 4.44//Previous Mean = 3.64), "developing ideas from a topic in speaking" (Mean = 4.31//Previous Mean = 3.50), "focusing on writing emails or letters in English" (Mean = 4.30//Previous Mean = 2.65), "listening to announcements or radio broadcasts in English" (Mean = 4.21//Previous Mean = 3.00), "listening to varieties of English accents" (Mean = 4.21//Previous Mean = 2.70), "making inferences from English reading texts" (Mean = 3.87//Previous Mean = 3.52). Moreover, there was a considerable rise in "metacognitive strategies" in the forms of "writing argumentative essays" (Mean = 4.16//Previous Mean = 3.01), "making plans before doing an English task" (Mean = 4.14//Previous Mean = 3.09), "making a point in English" (Mean =3.96//Previous Mean = 2.98), and "monitoring and evaluating an English task by myself" (Mean =3.71//Previous Mean = 2.98). Other strategies namely "compensation strategies" and "social strategies" almost remained stable.

Perceived Changes in the Time Devoted to the Use of Certain Learning Strategies

Table 6

Perceived Changes in the Time Devoted to the Use of Learning Strategies

Item Statistics				
My *time* devoted to	**Mean**	**Std. Deviation**	**N**	**Interpretation**
skimming for main ideas	3.9800	.69311	600	Increased slightly
scanning for specific details	3.5550	.73112	600	Increased slightly
memorizing new words/patterns	3.6467	.80596	600	Increased slightly
making predictions when carrying out a reading task	3.5267	.74176	600	Increased slightly
making inferences when reading a text	3.4600	.68983	600	Increased slightly
doing note-taking when listening to an English talk	4.1433	.75966	600	Increased slightly
summarizing key points after reading a passage	3.7033	.82652	600	Increased slightly
co-operating with peers when working in pairs/groups	4.0900	.88687	600	Increased slightly
self-evaluating after completing a task	3.7733	.73224	600	Increased slightly
expressing opinions in English	4.0417	.82729	600	Increased slightly

(*) Interpretation was based on *scale of change over time: Decreased greatly (1), Decreased slightly (2), Remained the same (3), Increased slightly (4), and Increased greatly (5)*

Table 6 features slight shifts concerning students' time devoted to the use of specific learning strategies for a specific skill compared to others after the introduction of VSTEP.3-5 test. An important pattern arises in the growing amount of time dedicated to active engagement and collaborative techniques. Specifically, the time spent on "doing note-taking when listening to an English talk" reached its peak with the highest Mean value of 4.14. Similar increases in time investment were also found in such learning strategies as "co-operating with peers when working in pairs/groups" (Mean = 4.09), "expressing opinions in English" (Mean = 4.04), "skimming for main ideas" (3.98). Another highlight of the table is the fact that students were inclined to spend time "self-evaluating after completing a task" (Mean =3.77), "summarizing key points after reading a passage" (Mean =3.70). By comparison, the amount of time devoted to "memorizing new words/patterns", "making predictions when carrying out a reading task", "making inferences when reading a text" almost remained unchanged (Mean = 3.36), with the Mean values equivalent to 3.64, 3.52, and 3.46, respectively.

Table 7

Students' Attitudes towards the Washback Effects of VSTEP.3-5 on Student Learning Strategies

Item Statistics				
The VSTEP.3-5 test has *strong* impact on how I use	Mean	Std. Deviation	N	Interpretation
Vocabulary learning strategies	4.078	.74820	600	Agree
Strategies for learning listening skill	3.7000	.89890	600	Agree
Strategies for learning reading skill	4.2167	.52617	600	Strongly agree
Strategies for learning speaking skill	4.4883	.61955	600	Strongly agree
Strategies for learning writing skill	4.4833	.58904	600	Strongly agree
Test-taking strategies	4.4433	.62778	600	Strongly agree
The VSTEP.3-5 test has *positive* impact on how I use				
Vocabulary learning strategies	3.9150	.84595	600	Agree
Strategies for learning listening skill	3.6100	.79306	600	Agree
Strategies for learning reading skill	3.9283	.66126	600	Agree
Strategies for learning speaking skill	4.3200	.67208	600	Strongly agree
Strategies for learning writing skill	4.2317	.64707	600	Strongly agree
Test-taking strategies	4.2833	.65608	600	Strongly agree

(*) Interpretation was based on *scale of agreement: Strongly disagree (1), Disagree (2), Neither agree nor disagree (3), Agree (4), and Strongly agree (5);*

Table 7 illustrates students' attitudes towards the effects of VSTEP.3-5 on their learning strategies in terms of two major dimensions of washback: intensity and direction. Regarding washback intensity, most of the students strongly agreed that "VSTEP.3-5 test has strong impact on how they use strategies," especially "strategies for learning speaking skill," with the highest Mean value (Mean = 4.4883), closely followed by "strategies for learning writing skill" (Mean = 4.4833) and "test-taking strategies" (Mean = 4.44). Other learning strategies, namely "Strategies for learning reading skill", "Vocabulary learning strategies," and "Strategies for learning listening skill," ranked the fourth (Mean =4.21), the fifth (Mean=4.07), and the last

(Mean=3.70), respectively. As for the direction of washback, there is a consensus among students on the notion that VSTEP.3-5 test has positive impact on how students use learning strategies. Again, three top strategies are "strategies for learning speaking skill" (Mean=4.32), "test-taking strategies" (Mean=4.28), and "strategies for learning writing skill" (Mean=4.23), sorted in order of Mean values. "Strategies for learning reading skill" and "vocabulary learning strategies" nearly shared the same ranking, with Mean values of 3.92 and 3.91, respectively. The lowest Mean value belongs to "Strategies for learning listening skill" (Mean = 3.61).

Findings from qualitative data

The results obtained from the surveys were clarified and elaborated by the data derived from the semi-structured interviews. To facilitate comprehension of the data demonstration, it was classified as "Before VSTEP.3-5" and "After VSTEP.3-5," and accompanied by relevant themes, codes, and highlighting examples.

Before VSTEP.3-5

Eight out of twelve respondents reported that before they learned about VSTEP.3-5 test, their most frequently employed strategies in learning all skills were "memory strategies". For example, in Listening skill, Student 06 exploited listening and dictation as her main strategy, whereas Student 02 and Student 05 preferred listening for the sake of doing tests only.

> "As for listening, I was very...of all skills, I think I was weakest in the listening skill, which means I think there was an effective strategy for me, which was dictation. I would listen and dictate so I could better understand the…the content of the audio file I am listening to." [S06]

> "For example, when it comes to listening skill, in the past, I often learned listening skills at lower levels of education, but it was usually test-oriented listening, that is, listening for the purpose of taking a test." [S02]

> "About listening, I would mainly practise and do listening exercises on the Internet." [S05]

In Reading skill, Student 01 and Student 12 mainly "practiced reading tests and exercises on the websites." Concerning learning vocabulary, Student 04 and Student 02 "learned by memorizing and repetition". While Student 04 "acquired new words from reading passages and memorized them," Student

02 learned "vocabulary occurring in tests or exams." Specifically, "translation strategies" were employed in Student 07's way of learning new words. She would look up the meaning of new words in a dictionary in Vietnamese, or translate difficult words into Vietnamese to better understand the meaning of new words. She also made use of a notebook to list and categorize new words.

> "About my learning strategy, for learning new words, I had a small notebook to write down the vocabulary I often encountered…frequently encountered vocabulary, or new vocabulary…I would write down in a list, in a notebook…I would usually divide it into three columns: a vocabulary column, a column of examples, and the other column is…the column of meanings. I could take notes of the meaning in both English or Vietnamese. For the words that I found difficult to understand, I would look them up in the dictionary and translate them into Vietnamese so that I can understand and grasp them better." [S07]

In Writing skill, Student 12 reflected on "learning from Writing samples on the Internet." Added to this, Student 01 described his strategy as "learning by picking up and memorizing phrases from other resources," whereas Student 11 "had no specific strategies at all." These findings were consistent with that of the questionnaires.

After VSTEP.3-5

After acknowledging VSTEP.3-5 test, Student 04 claimed that she learned new words by "brainstorming words in context," while Student 08 admitted that VSTEP.3-5 test made her realize there was an urgent need to learn more about collocations and related words.

> "[…] For example, that word can have many meanings, but when I apply it in context, it will have different meanings, so like... I will try to brainstorm that vocabulary in that context […] have different meanings." [S04]

> "After learning about VSTEP level 3-5, I feel that I need to learn more vocabulary as well as collocations that belong to the higher band so that I can achieve the score…the maximum I want." [S08]

Besides, regarding Reading skill, specific strategies were identified by participants. Several highlighting codes were reported such as "learning from peers" (Student 03), "learning synonyms from the reading passages" (Student 04), or "reading newspapers on websites" (Student 01).

> "Yes, actually, there have been a few adjustments…in my strategy…Because I see that in the VSTEP.3-5 test, especially reading, there are often many synonyms. Before that, I used to learn separate single word and after this […] I would learn...try to learn. For example, if I learn that vocabulary, I will find and learn more synonyms of that vocabulary." [S04]

As for the effect of VSTEP.3-5 on listening strategies, while Student 04 seemed to focus more on note-taking and inferring strategies, Student 03 emphasized daily practicing routines and listening to varieties of English accents. She maintained her overall strategies, yet seemed to pay more attention to listening to lectures and VSTEP-related themes and topics.

> "As for listening...listening skill, I think it's (the test) a multiple-choice test, it's very difficult. Because it "puzzles" in a way that we have to understand…We have to have knowledge about that topic, then we have to understand that…for example, initially, they say something differently, then after that they will say it differently, with the answer…the answer can't be... it can't just be the straightforward words that you hear, but the answer is…*We have to deduce it by ourselves from those things. When I learn to listen, I also learn the "note-taking" strategy, so I find that strategy very useful.* Because…normally, if I just listen without writing down notes, I will easily forget and not concentrate, so after learning to listen in class, I feel like... Every time I listen, I will note down the main ideas." [S04]

> "I think that in order to develop my listening skills, I have to listen to a lot of English every day, and also a lot of the intonation, and different regional accents such as British accent, Australian accent, American voice. [...] Currently, my learning strategy remains basically the same. But I focus on listening more to the lectures. Because I feel like in the lecture part in VSTEP, not only do I have to hear the words, but I also have to have an understanding of the lectures they teach. So, I'm trying to expand into as many topics as possible." [S03]

Notable changes were reported in students' choice of strategies for learning productive skills. The most obvious effect of VSTEP.3-5 test on Student 06's use of speaking strategies was the shift towards "cognitive and social strategies" through "practicing developing topic and ideas" and "more interaction with lecturers for receiving feedback and correction" (S06).

"Cognitive and social strategies" continued their patterns in students' choices for learning writing skill. Student 01 confirmed her "enhancement in academic vocabulary," indicating a more profound involvement with language, while Student 06 demonstrated an increased openness to "learning

from comments and feedback" provided by her peers and her teachers, suggesting a collaborative approach to improving her skill. Student 12 added that she paid greater attention to "writing emails and letters" which demonstrates the practical utilization of writing abilities in real-life situations, since it was a component of VSTEP.3-5 test design. This data highlighted the effect of VSTEP.3-5 in promoting holistic development in students' writing skills.

Exceptionally, Student 04 asserted that each individual might have a different learning strategy, yet she held a firm belief that the VSTEP.3-5 test's objectives matched well with her acquired strategies at university.

> "I think whether the learning strategy can be the same or not depends on each person, meaning each student will surely have a different strategy. But for me, the exam objectives are related to the strategies I acquired at university." [S04]

However, besides those most preferred strategies above, two respondents were strongly in support of "test-taking strategies" in their preparation for VSTEP.3-5. For instance, Student 02 put more focus on listening to academic materials and the format of the exam instead of listening extensively like before.

> "[…] For example, before (VSTEP.3-5), I listened in a more comfortable way, listening to materials called "extensive listening", such as listening to music, or radio…but now I think I should...if I put my academic goals as my priority, I should emphasize the resources I exploit. For example, in order to pass the exam, I think I will promote the practice of listening to academic documents rather than listening to music or listening to the radio." [S02]

Similarly, Student 06 emphasized the importance of practicing according to the test format and test-oriented materials.

> "[…] I think I will study that format (VSTEP.3-5) carefully, and I will focus on studying according to the formats, meaning the types of questions that are required and…and follow the test format more closely…so that I can just...that means I can get familiar with the test, and later be used to the test, so I can take the exam to save time for myself. Because I think that when I get used to the format, I will save a lot of time." [S06]

Discussion

The findings demonstrated a relatively significant shift in undergraduate students' utilization of language learning strategies, following the implementation of VSTEP.3-5 test. In fact, prior to VSTEP.3-5 test, it is worth mentioning that the most commonly used strategies belonged to the group of "compensation strategies" and "memory strategies" as described in Oxford's (1990) taxonomy, such as "rephrasing words/phrases/sentences," "inferring the meaning of words in context," "listening for dictation," "memorizing words or phrases," "rote-learning from Writing samples on the Internet." After the introduction of the VSTEP.3-5 test, there was a remarkable change in how learning strategies were distributed and used. The use of "cognitive and metacognitive strategies" among students increased, which is particularly noteworthy. The increasing patterns identified in generating ideas from topic sentences in both written and spoken forms, along with the emphasis on specific activities such as "composing emails or letters in English" and "comprehending different accents," emphasized an increased cognitive engagement in the processes of language acquisition. Furthermore, the increase in "metacognitive strategies," such as pre-planning English activities and self-monitoring, highlights a developing understanding and mastery of the learning process. It is also crucial to note that there was an increasing tendency among students to dedicate time to metacognitive processes, such as self-assessment. In contrast, certain approaches, such as rote memorizing of vocabulary and employing predictive or inferential skills while reading, have demonstrated little alterations in the amount of time required. This could suggest a consistent dependence on these methods or a perceived decrease in importance as compared to other techniques in the context of language learning after the implementation of the VSTEP.3-5 test. This finding confirms Kong and Zhang's (2024) viewpoint related to HSK washback. In their study, positive washback was reflected through the fact that the test (HSK) "enabled students to adjust their learning strategies appropriately."

Also, the change towards more intentional and thoughtful ways of students' language learning processes could be an indicator of "positive washback," based on Bailey's (1996) classification of washback dimensions. Most importantly, contrary to some past studies in which negative washback of a high-stakes test outweighed positive washback in terms of learning strategies (Aftab et al., 2014; Cheng et al., 2011; Sukyadi & Mardiani, 2011), the findings appeared to be aligned with the latest studies conducted by Fan and Yang (2020), Chak (2023), which highlighted the use of cognitive and metacognitive strategies as the primary choices, rather than test-related strategies.

Nonetheless, it is indisputable that there are still several students who advocated the importance of "test-taking strategies" when preparing for the VSTEP.3-5 test. In order to achieve high results on the test, they believed that one needed engagement in intensive practice using test-related materials and test-taking strategies. Washback is therefore considered "negative" at this stage, based on Bachman and Palmer's (1996) perspective on the direction of washback. This phenomenon is explicable and consistent with the existing body of literature. According to Green (2007), the design of a test can tempt language learners to focus on improving their test-taking tactics rather than establishing long-term learning skills, which is especially true in Asian nations where success in high-stakes exams can have a significant impact on the test-takers' future in multiple areas.

Overall, the results of the present study revealed a considerable positive change in the majority of students' use of language strategies after the introduction of VSTEP.3-5 test. These findings may appear paradoxical in light of previous washback studies a decade ago, yet in line with those of recent washback research on language learning strategies.

Conclusion and Implications

In conclusion, these findings highlighted the ever-changing characteristics of language learning processes and the subtle adjustments made by students in response to the washback effects of VSTEP.3-5 test. By adjusting how they managed their time, students prioritized tactics that encourage active engagement, cooperation, and metacognitive reflection. This might help them improve their language proficiency and comprehension abilities, in accordance with the requirements of the VSTEP.3-5 test and beyond.

There are certain implications worthy of educators' attention. Through acknowledging washback effects of VSTEP.3-5 test on students' learning strategies, teachers and instructors can better tailor their teaching strategies to align with the changing needs and preferences of students, eventually promoting more efficient and significant language learning experiences.

Limitations and Recommendations

First and foremost, regarding participants and research site, the study was restricted to student participants of a single representative university in Central Vietnam. The generalizability of the study might have been enhanced if it had included a larger sample size and the participation of additional universities across the country. Moreover, given that the nature of washback

study is of longitudinal type, it would have been more advantageous if the study had been conducted throughout an entire academic year rather than just a semester. It is recommended that further studies be implemented including a larger sample size and a longer duration.

Acknowledgements

This study was funded by "Chulalongkorn University's Graduate Scholarship Program for ASEAN or Non-ASEAN Countries."

About the Authors

Phuong Ai Ngoc Doan: A full-time lecturer of English at Faculty of English, Hue University of Foreign Languages and International Studies (HUFLIS), Hue City, Vietnam. She is currently pursuing her Ph.D. in EIL (English as an International Language) Program at Chulalongkorn University, Bangkok, Thailand. Her research areas of interest include TESOL, English Language Methodology, and Language Assessment and Evaluation.

Chatraporn Piamsai: An Assistant Professor and a full-time lecturer at Chulalongkorn University Language Institute (CULI) and the English as an International Language (EIL) program, Chulalongkorn University, Bangkok, Thailand. Her research interests involve various aspects of language assessment, including English proficiency and English as a Specific Purposes test development, alternative assessment and language assessment in a local context.

References

Aftab, A., Qureshi, S., & William, I. (2014). Investigating the washback effect of the Pakistani intermediate English examination. *International Journal of English and Literature*, *5*(7), 149-154. https://doi.org/10.5897/IJEL2013.0521

Ahmed, A. M. (2018) *Washback: Examining English language teaching and learning in Libyan secondary school education.* [Doctoral thesis, University of Huddersfield, United Kingdom]. Retrieved from http://eprints.hud.ac.uk/id/eprint/34799/

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

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing, 13*(3), 257-279. https://doi.org/10.1177%2F026553229601300303

Bui, T. H. (2018). *Washback of the proficiency assessment as exit requirement on non-English major students at a law school.* [Bachelor's Graduation Paper, Vietnam National University]. Retrieved from https://www.scribd.com/document/479562201/Bui-Thanh-Huong

Chak, M.-C. (2023). Washback in language learning strategies under high stakes language testing - A study of the Hong Kong secondary system. *rEFLections*, *31*(1), 1–24. https://doi.org/10.61508/refl.v31i1.269539

Chen, L. (2002). *Taiwanese junior high school English teachers' perceptions of the washback effect of the basic competence test in English*. [Doctoral dissertation, Ohio State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=osu1039116364

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%2F0265532210384253

Cohen, A. D., & Chi, J. C. (2004). Language strategy use inventory. In R. M. Paige, A. D. Cohen, B. Kapper, J. C. Chi, & J. P. Lassegard (Eds.), *Maximizing study abroad: A students' guide to strategies for language and culture learning and use* (pp. 16-22). Center for Advanced Research on Language Acquisition, University of Minnesota.

Dunlea, J., Spiby, R., Nguyen, T. N. Q., Nguyen, T. Q. Y., Nguyen, T. M. H., Nguyen, T. P. T., Thai, H. L. T., & Sao, B. T. (2018). *APTIS-VSTEP comparability study: Investigating the usage of two EFL tests in the context of higher education in Vietnam* (British Council Validation Series Report No. VS/2018/001). British Council. https://www.britishcouncil.org/sites/default/files/aptis-vstep_study.pdf

Fan, H., & Yang, M. (2020). The washback effects of English tests on language learning strategies: A comparative case study based on the data analysis of the English and non-English majors in China, *International Conference on Computer Information & Big Data Applications (CIBDA)*, 41-47. http://dx.doi.org/10.1109/CIBDA50819.2020.00018

Green, A. (2007). *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

Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge University Press.

Jamalifar, G., Salehi, H., Tabatabaei, O., & Jafarigohar, M. (2021). Washback effect of the English proficiency test (EPT) on PhD candidates' language learning strategies. *Journal of Language and Translation, 11*(3), 185-198. https://www.doi.org/10.30495/ttlt.2021.683485

Kong, F., & Zhang, Y. (2024). Investigating the washback of the Chinese language proficiency test (HSK): From the perspective of CSL students. *Sage Open, 14*(1) https://journals.sagepub.com/doi/epub/10.1177/21582440231224599

Nguyen, H., & Gu, Y. (2020). Impact of TOEIC listening and reading as a university exit test in Vietnam. *Language Assessment Quarterly, 17*(2),147-167. https://doi.org/10.1080/15434303.2020.1722672

Nguyen, T. L. (2017). Washback effects of VSTEP on teaching English at University of Languages and International Studies – Vietnam National University. *VNU Journal of Foreign Studies, 33*(4). https://doi.org/10.25073/2525-2445/vnufs.4177

Nguyen, T. N. Q. (2018). A study on the validity of VSTEP writing tests for the sake of regional and international integration. *VNU Journal Of Foreign Studies, 34*(4). https://doi.org/10.25073/2525-2445/vnufs.4285

Nguyen, T. N. Q. (2020). Vietnamese standardised rest of English proficiency: A panorama. In L.I.-Wen Su, C. Weir, & J. Wu (Eds), *English language proficiency testing in Asia: A new paradigm bridging global and local context* (pp. 71-100). Routledge.

Oxford, R. (1990). *Language learning strategies: What every teacher should know*. Newbury House.

Pan, Y.-C. (2008). A critical review of five language washback studies from 1995-2007: Methodological considerations. *Shiken: JALT Testing & Evaluation SIG Newsletter, 12*(2), 2-16.

Rahimi, F., Esfandiari, M. R., & Amini, M. (2016). An Overview of Studies Conducted on Washback, Impact and Validity. *Studies in Literature and Language, 13* (4), 6-14. http://dx.doi.org/10.3968/8877

Sadeghi, K., Ballidag, A., & Mede, E. (2021). The washback effect of TOEFL iBT and a local English proficiency exam on students' motivation, autonomy and language learning strategies. *Heliyon, 7*(10), 1-7. https://doi.org/10.1016/j.heliyon.2021.e08135

Sukyadi, D., & Mardiani, R. (2011). The washback effect of the English national examination (ENE) on English teachers' classroom teaching and students' learning. *K@ ta: A Biannual Publication on the Study of Language and Literature*, *13*(1), 96-111. https://doi.org/10.9744/kata.13.1.96-111

Vietnam's Ministry of Education and Training. (2015). *Government decision 729: Issuance of the test of English proficiency levels 3–5 in the Vietnam's six-level framework of foreign language competency* (pp.1-4). National Foreign Language Project 2020.

Xu, Q., & Liu, J. (2018). *A study on the washback effects of the test for English majors (TEM).* Springer Nature Singapore Pte Ltd.

Appendix

VSTEP.3-5 Test Specifications

Skill (s)	Test Constructs	Timing	Text types	Test tasks/ response formats
Listening	Sub-skills of listening from level 3 to level 5: listening for details, listening for gist, identifying speakers' purposes, opinions, attitudes making inferences based on information available from the listening texts.	40 minutes, including transfer time	**Part 1** (10-11 minutes): 8 spoken announcements, instructions, warnings etc. (30-60 words/ spoken text at the rate of 130-170 words/ minute) **Part 2** (10-12 minutes): 3 conversations between two fluent speakers (in which varieties of English may appear) with the speed of 110-170 words/ minute) **Part 3** (12-15 minutes): 3 excerpts of lectures/ presentations/ long texts given by native speakers/ ESL speakers.	**Part 1:** 8 multiple choice questions with 4 options A, B, C and D. **Part 2**: 12 multiple choice questions with 4 options A, B, C and D. **Part 3:** 15 multiple choice questions with 4 options A, B, C and D.
Reading	Sub-skills of reading from level 3 to level 5: Scanning: identifying,	60 minutes, including transfer time	**Part 1**: 1 text at Level 3 about familiar topics (450 words per text)	10 multiple choice questions with 4 options A, B, C and D for each Part.

	comparing and transferring specific details Skimming: Identifying the main points and key conclusions in clearly-written texts. Paraphrasing, guessing vocabulary in context Identifying types of text Reading for comprehension and make inferences based on information available in texts.		**Part 2**: 1 text at Level 4 about Sciences or Social Sciences (450 words per text), mostly containing explicit information. **Part 3**: 1 text at Level 4 about Sciences or Social Sciences or other specialized areas, containing both explicit and implicit information (450 words per text). **Part 4**: 1 highly specialized text at Level 5, containing both explicit and implicit information (500 words per text)	
Speaking	Sub-skills of speaking from level 3 to level 5: Responding to interview questions, describing, explaining a specific situation/experience/expressing feelings etc. (Part 1) Problem-solving skill, presenting an argument, negotiating etc. (Part 2) Public speaking skill, developing and expanding arguments, giving examples, persuading, responding to in-depth questions, expressing opinions (Part 3)	12 minutes	**Part 1** (3 minutes) Social interaction **Part 2** (4 minutes, including 1 minute for preparation) Solution Discussion **Part 3:** (5 minutes, including 1 minute for preparation) Topic Development	**Part 1:** 4-5 short interview questions about 2 random topics. **Part 2:** A situation in which the candidate is given 3 options and required to choose the best one with justification. **Part 3**: A monologue in which the candidate is given a prompt (1 topic sentence and 3 expanding ideas) and required to develop the topic, using his/her reasons, opinions and experiences.
Writing	Sub-skills of writing from level 3 to level 5:	60 minutes	**Part 1** (20 minutes): Emails, letters (informal or formal)	**Task 1**: Writing an email/ letter of at least 120 words

	The ability to carry out such speech acts as: thanking, apologizing, suggesting, inviting, etc. The ability to produce personal letters asking for information or describing personal experiences/ feelings or for communicative purposes. The ability to present ideas about concrete as well as abstract topics. The ability to give arguments/ opinions/ reasons and include any relevant examples from one's own knowledge or experience.	**Part 2** (40 minutes): Argumentative essays	in response to a given email/letter (of 80-90 words) **Task 2**: Writing an essay of at least 250 words to discuss the given prompt (e.g. a 50-60 word text about a controversial topic related to Vietnam/ other countries in the region).

(Vietnam MOET's Decision 729, 2015, pp.3)