

CHAPTER III

RESEARCH METHOD

This chapter is organized into distinct sections. Initially, the section revisits the fundamental research questions, setting the stage for the subsequent discussions. The next section outlines the chosen research design, and following this, it delves into the research procedure, then other two sections elaborate the research site and participants, crucial components of the study. Then, it discusses ethical considerations, followed by a comprehensive exploration of data collection methods and instrumentation. The other subsequent sections, are dedicated to data analysis and considerations of validity, reliability, trustworthiness, and transferability, offering a holistic view of the research methodology. The chapter concludes with a final section, summarizing key findings and insights gleaned from the preceding discussions.

3.1 Research Questions

This research investigated three research questions, as outlined in the first chapter:

- 1) Why does the vocational higher institution adopt the TOEIC® as an English exit test?
- 2) How does the TOEIC® affect ESP teachers' teaching practices?
- 3) How does the TOEIC® affect students' learning practices of English?

3.2 Research Design

This study is grounded in pragmatism and transformative philosophy (Mertens, 2007), since it tries to accommodate the most effective way of addressing the research questions. For that purpose, mixed-method research (MMR) was adopted. This study focuses on stakeholders' perceptions of the impact of the TOEIC test as

an English exit test, and it is based on an explanatory sequential mixed methods design that incorporates elements from both quantitative and qualitative traditions. In MMR, Johnson (2012) proposed what was coined "dialectical pluralism," in which a meta paradigm with a dialectical approach listens to each research topic and aims accordingly, allowing researchers to include insights from opposing paradigms (Johnson, 2012). Combining several methodologies yields a more comprehensive knowledge of study challenges than each strategy alone does (Bergman, 2008). Furthermore, mixed methods research is the culmination of the "evolution" of single-method research (Tashakkori & Teddlie, 2003). The practical contribution of this study is to enhance the positive impact of TOEIC® as an EET and to promote more alignment among macro, meso, and micro level stakeholders in policy enactment.

For this study, an explanatory sequential design (QUAN → qual) is adopted from among three distinct mixed methods designs since it is a good fit for addressing the provided research issues (Creswell & Plano Clark, 2018). The notation shows that the design puts greater emphasis on quantitative data, with an additional qualitative component. Prior to making the final study results, supplemental data gathering and analysis procedures are utilized to analyse an unexpected, unexplained, or unusual finding (Morse et al., 2006). If the quantitative sample is no longer available, the researcher must create a new sample based on the qualitative criterion of 'good informants' (Spradley, 2016). These qualitative findings are then included in the quantitative results. Thus, this study was descriptive in nature, with both quantitative and qualitative features with explorative case study design (Creswell, 2014; Hamied & Malik, 2017).

The collection and analysis of quantitative data on the specified topics is the first step in the explanatory sequential design. The quantitative data finding is followed up by qualitative data-collecting via interviews to determine how and why the phenomenon happens. To summarize, the design allows quantitative approaches from stage one to be explained by qualitative methods from stage two in order to

add depth and gain a more comprehensive understanding of stakeholders' viewpoints on the TOEIC® washback (Creswell & Plano Clark, 2018; Creswell, 2014). Flyvbjerg (2011) believed that these methods can complement with one another:

...if you want to understand a phenomenon in any degree of thoroughness ..., you need to do case studies. If you want to understand how widespread the phenomenon is, how it correlates with other phenomena and varies across different population, and at what level of statistical significance, then you have to do statistical studies. If you want to understand both, which is advisable if you would like to speak with weight about the phenomenon at hand, then you need to do both case studies and statistical analyses. The complementarity of the two methods is that simple, and that beautiful (p. 314).

In the triangulation design, the researcher uses a range of sources, methodologies, investigators, and ideas to provide supporting evidence (Creswell & Creswell, 2013). While triangulation is frequently used to gather converging information, it can also be used to determine different interpretations or uncover inconsistencies and contradictions, which can be useful for refining frameworks and conclusions (Duff, 2008). In the current work, data and methodological triangulation were used to better understand the intricacy of the test washback.

The proposal for this study was initially prepared during the COVID-19 pandemic, when the Community Activities Restrictions Enforcement (PPKM) was enforced. While the study sought to analyse institutional language policy, engaging numerous stakeholders, some targeted institutions chose not to participate. As a result, the study had to confine its scope to a single research site. At this particular research site, there was a restricted number of participants, comprising policymakers and ESP teachers. As a result, doing quantitative studies with a limited sample size is impractical. Thus, a mixed-methods design was used to address the third research question for student participants, while the first and second research questions were answered using a case study. Currently, washback

studies in one research location were prevalent in previous washback studies (Allen, 2016; Hung & Huang, 2019; Pan, 2022; Saglam, 2018). Through a deeper exploration of the policies, practices, and perspectives within a particular institution, this study aimed to provide a detailed analysis, providing a rich source of insight.

Table 3.1 Overall Research Design

Phase 1		
		Student Questionnaire
Phase 2		
Policymaker Interviews	Teacher Interviews	Student Interview

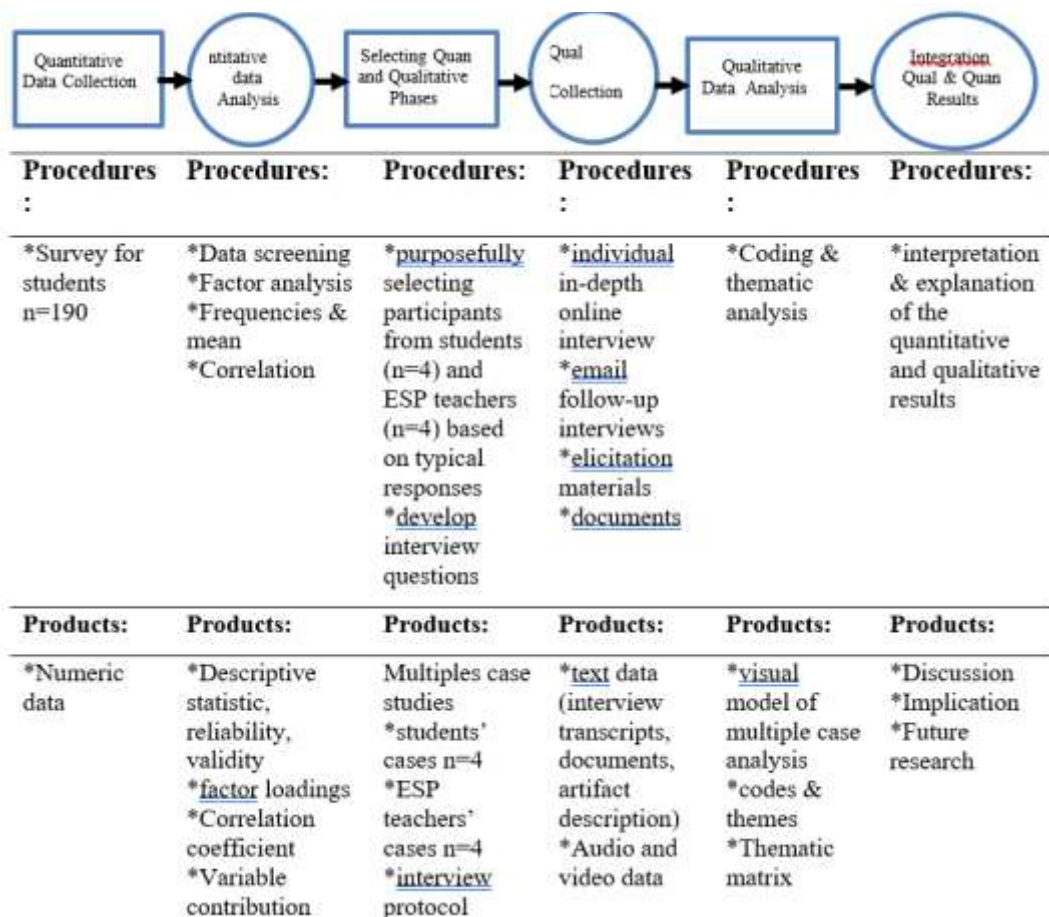
As seen from the **Table 3.1**, the first phase involved quantitative data collection through distributed questionnaires for student, which were analysed using statistical tools such as SPSS 20 and JASP. The subsequent phase involved qualitative data collection through interviews to illuminate the quantitative findings, with participants from each stakeholder group undergoing focused interviews. This integrative approach facilitated answers to the primary research questions by linking qualitative insights with quantitative findings, providing a thorough understanding of the research problem. The choice of a mixed methods design stemmed from both intrinsic and strategic considerations, combining the strengths of quantitative and qualitative methods for a nuanced understanding (Hamied & Malik, 2017), and to refine the quantitative results (Fraenkel et al., 2018). This approach, which addresses the ongoing debate between quantitative and qualitative approaches, enhances the quality of the research, increases the chances of acceptance for publication, and contributes valuable data for practical and policy-oriented outcomes.

3.3 Research Procedure

Having received the ethical clearance, the researcher officially proposed the targeted institutions to access data. Then, the researcher needed to administer the

questionnaire adapted from previous studies. Instrument administration went through several steps: drafting, trialling, piloting, and questionnaire revision. The final draft was distributed to a small sample for checking validity and reliability. After ensuring that the questionnaire has appropriate validity and reliability, it was distributed to the targeted research participants, i.e., students. Following that, she began purposively selecting participants for semi-structured interviews with accessible Heads of Departments, ESP teachers, and students. Data analysis was performed by interpreting and combining the results for meaningful findings. As the research adopted an explanatory sequential design, it followed specific procedures (Creswell & Plano Clark, 2018). Here is the diagram of the research procedure adapted from (Creswell & Plano Clark, 2018; Ivankova et al., 2006).

Figure 3.1 Research Procedure



The data collection process involved two distinct phases: quantitative and qualitative. In response to the global pandemic, the quantitative phase employed an online questionnaire. Following that, the qualitative phase with semi-structured interviews with department heads, ESP teachers and students were conducted. The research design comprehensively addressed both the macro level, involving institutional decision makers, and the micro level, involving ESP teachers and students. To this end, a variety of data collection methods were used to capture different stakeholder perspectives. Realizing that test washback is a complex phenomenon, the use of a range of data collection methods, as exemplified in this research, is expected to provide a comprehensive and insightful picture.

3.4 Research Site

The study was conducted at a vocational tourism college in Jakarta, which has long implemented TOEIC as an English exit test policy. Then, it was chosen for its accessibility due to the researcher's five-year tenure as a faculty member, which ensured the study's ease of implementation. A more naturalistic approach was enabled by familiarity with the location, participants, and stakeholders, in line with the goal of capturing data as it unfolded in its authentic context. The decision was made to collect as much data as possible. Site selection required careful consideration of factors such as relevance to the research questions, access to the target population, and the site's ability to meet specific requirements to ensure the study's robustness.

3.5 Research Context

The Test of English for International Communication (TOEIC) was developed by the Educational Testing Service (ETS), a non-profit organisation based in New Jersey, USA, with offices around the world, including Indonesia. Beyond its primary purpose, TOEIC serves a variety of secondary purposes, including verifying proficiency, qualifying for positions, enhancing credentials, monitoring progress, and setting learning goals (Educational Testing Service.,

2015a, 2016). These uses include crucial aspects such as hiring, placement, promotion of candidates, measuring language proficiency, and evaluating English progress (Educational Testing Service., 2015a, 2016).

Initially designed to assess English proficiency in international business scenarios, a revised version was introduced in 2006, targeting individuals who use English in global work contexts (ETS, 2013; Educational Testing Service., 2015b, 2016; Powers & Powers, 2015). Beyond its primary function, TOEIC fulfils various secondary roles such as confirming proficiency, qualifying for positions, enhancing credentials, monitoring progress, and setting learning goals (Educational Testing Service., 2015a, 2016). These applications include vital elements such as hiring, placement, candidate promotion, language proficiency measurement, and English progress assessment (Educational Testing Service., 2015a, 2016). The latter roles are relevant with the goals of VHE institutions which were developed against the background of industry's need for skilled labour (Finlay in Suharno et al., 2020). Thus, VHE aimed to contribute to a country's economic growth through labour independence.

3.6 Research Participants

This study is directed at all VHE institutions which employ TOEIC® as an English exit test, with a special emphasis on tourism VHE, in response to the growing demand for English skills in the tourism employment. Unfortunately, national data on the use of English Exit Tests (EETs) in VHE institutions are lacking. As a result, the study chose a private tourist VHE institution through convenience sampling, emphasizing voluntary participation and accessibility. The study's participants include policymakers such as department heads, one test coordinator, and one language coordinator, as well as ESP teachers and students studying hotel management, travel, hospitality, and tourism. Furthermore, HRDs and General Managers from the hotel and travel industries were interviewed to learn about their perspectives on TOEIC certification. The student participants were from three departments (Hotel management, travel, and tourism and hospitality) with a total of 450 and a sample of

190 participants (Hair et al., 2018), with pilot study participants also included as in **Table 3.2**.

Table 3.2 Participants in Pilot and Main Study

Types of participants	Students	ESP Teachers	 Policymakers
First piloting	5	2	1
Second piloting	30	3	1
Main study	190	6	4
Total Participants	200 participants		

The draft instrument went through a piloting phase as an essential component of the development process prior to the official piloting of the questionnaire. Students, ESP teachers, and department heads participated in this phase and offered input on the appropriateness, representativeness, and relevancy of the content. Three Indonesian English teachers and two PhD candidates from the UK and Australia were given the TOEIC to simulate real-world situations. The participants answered the questions as though they were professors at colleges that required the test to be completed in order to graduate. Response times from the group were noted, and comments on structure, substance, and clarity were gathered. Improvements were made to the format, instructions' clarity, and the language of the questions. There were three participant groups in the study: students, ESP teachers, and policy makers. Each group had a separate set of research questions.

3.6.1 Policymakers

To address the first research question, there were four participants who had reported their consent. As seen from **Table 3.3**, they comprised of head of departments of tour and travel and hotel management, one served as testing coordinator, and one language coordinator. As part of research ethic, the participants' identities were anonymous, and replaced by P1, P2, P3, and P4. In total, there were four participants that were contacted and agreed to participate by assigning the letter of consent. The interviews were conducted in their most convenient time, and last for

30-60 minutes. Follow up interviews were conducted as to clarify specific details. The demographic data of the participants can be seen in **Table 3.3**.

Table 3.3 Policymakers Profile

Participants	P1	P2	P3	P4
Gender	Male	Female	Female	Female
Age	55	42	53	57
Education	S2	S3	S2	S2
Industry experience	Yes	Yes	Yes	Yes
Expertise	Travel and tour	tourism management	tourism management	Tourism destination
Role	HoD	HoD	Testing coordinator	Program Developer

The majority of the research participants, predominantly female, with one male head of department, have industrial experience that enriches their teaching practice in this institutional context. All participants have roles in structural and academic professions that include teaching, research and community service obligations. Their ages range from 42 to 57, and each participant has served in the institution for more than fifteen years. P2 has a doctorate degree in service management, while the other two are pursuing doctorates in tourism and hospitality management. The interviews, which were recorded and transcribed, underwent thematic analysis to address the sub-research questions. The selection of participants aimed to include diverse perspectives of policymakers involved in the implementation of testing policy, in line with the research objectives, ethical considerations and participant characteristics (Latunde, 2017; Newington & Metcalfe, 2014; Saunders, 2012).

3.6.2 ESP Teachers

Having gained the institution consent, the researcher sought personal consent from the teacher participants. Referring to some methodological consideration in Chapter 3, the researcher adopted some potential teachers who were willing to participate in the study, with some consideration, as the following:

- Teachers who were willing to be interviewed: 6 participants
- Teachers who were willing to be observed for four sessions: 3 participants

- Teachers who agreed to be audio and video recorded: 3 participants
- Teachers who were willing to encourage their students to ignore the video camera: 3 participants
- Teachers who were ready to engage in discussion after the classes were over: 3 participants
- Teachers who were selected to be presented in teacher cases: 3 participants

Table 3.4 provided demographic profile of six teachers who were willing to be interviewed.

Table 3.4 Teacher' s participants' Profile for Interviews

No	Pseudonym	Gender	Teaching Experience	
			English	TOEIC
1	Awan	Male	> 15 years	> 6 years
2	Lia	Female	> 10 years	> 5 years
3	Atiyas	Female	> 20 years	> 6 years
4	Tere	Female	> 15 years	> 6 years
5	Risa	Female	> 20 years	> 6 years
6	Feri	Male	> 10 years	> 5 years

The table included both male and female teachers with Master's degrees and extensive experience teaching English, with a special emphasis on TOEIC education. Awan and Tere bring over 15 years of general English instruction and 6 years of TOEIC teaching experience to the table. Lia, Atiyas, and Risa each provide a plethora of TOEIC knowledge as well as more than 6 years of general English teaching experience. Feri has 7 years of general English teaching experience and 5 years of TOEIC teaching experience. Because of their considerable experience and significant qualifications, the teachers are well-prepared participants who can provide insightful viewpoints on the study's subject matter.

Table 3.5 Teacher's Participant Profiles for Case study

No	Pseudonym	Gender	Teaching Experience	
			English	TOEIC
1	Awan	Male	> 15 years	> 6 years

2	Lia	Female	> 10 years	> 5 years
3	Atiyas	Female	> 20 years	> 6 years

Among all the six participants to be interviewed, three teachers were willing to be observed, and recorded for further analysis.

Table 3.6 Teacher Participant for Classroom Observation

Pseudonym	Subject	Grades/credit	No of sessions observed	No of minutes observed
Tere	Business English	3 rd grade/3 sks (3 hours)	4	720 minutes
Tiyas	English for hotel	3 rd grade/2 sks (2 hours)	4	480 minutes
Lia	Business English	3 rd grade/3 sks (3 hours)	4	720 minutes

These three teacher-participants were willing to participate in all data collection processes, such as classroom observations, and interviews.

3.6.3 Students

Research question three deal with students' perspectives on the washback of TOEIC on their learning. The data collection was performed with survey and interviews as seen in the **Appendices 1-5**.

3.6.3.1 Participants for Survey

For this purpose, a total of 195 responses submitted the online survey, with only five disagreeing. As in **sub-chapter 3.7.3**, provides comprehensive discussion on the instruments. After ensuring that all responses were complete, 190 responses were available for the analysis. This represented approximately 38% of the total student population. The figures below portrayed the demographic data of the students-participants. Demographic data are crucial to identify the trends and correlations with other variables, such as participants' background.

3.6.3.1.1 Gender

The pie chart below depicted the detailed data in terms of respondents' gender.

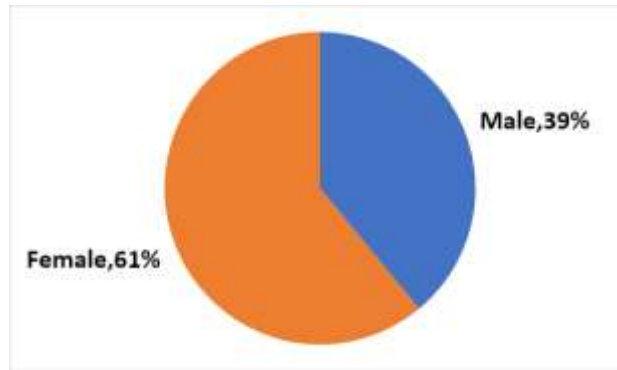
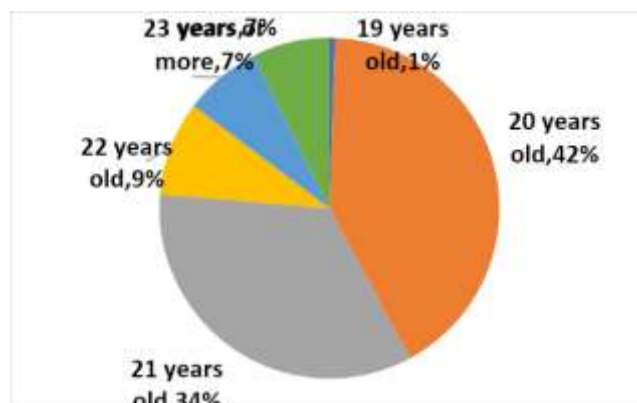


Figure 3.2 Gender

As seen from the figure, most respondents were female (61%), and the rest were male (39%). A total of 116 female students participated in the survey. However, only 74 male students participated in this study. The fact that the number of female students outweighed that of males was evident in the Higher Education Statistics issued by the Ministry of Research and Education (Kemdikbud, 2020). The statistics showed that the number of female students in the 2020 new intake in vocational education in DKI Jakarta was 10.350 for males and 13.105 for females (p.98). This fact was also evident in undergraduate academic program.

3.6.3.1.2 Age

Figure 3.3 displays the age range of the respondents, which ranged from 19 to older.

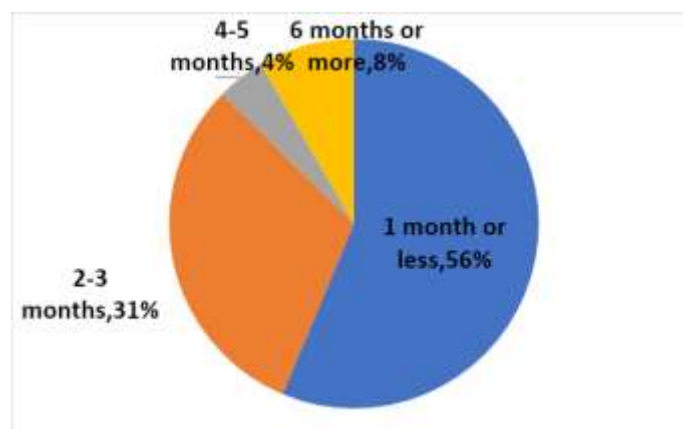


The majority of responders (42 %) had not yet reached the age of 20, while 34% were between the ages of 21 and 22. Furthermore, 9% of the respondents was 22

years old, with persons aged 23 years or older accounting for 7% of the total respondents. In contrast, 7% of respondents was under the age of 19.

3.6.3.1.3 Length of Test Preparation

This item aimed to portray students' learning time before they underwent the TOEIC test. The majority reported within one month or less as 56%, followed by 2-3 months (31%), 6 months or more (9%), and 4-5 months, 4%). The limited preparation time is due to their hectic learning schedule, either inside their campus or during job training.



3.6.3.1.4 English Teaching Experience

This section elaborates on respondents' English learning experiences through formal education. As a foreign language, English teaching in formal education varies across places.

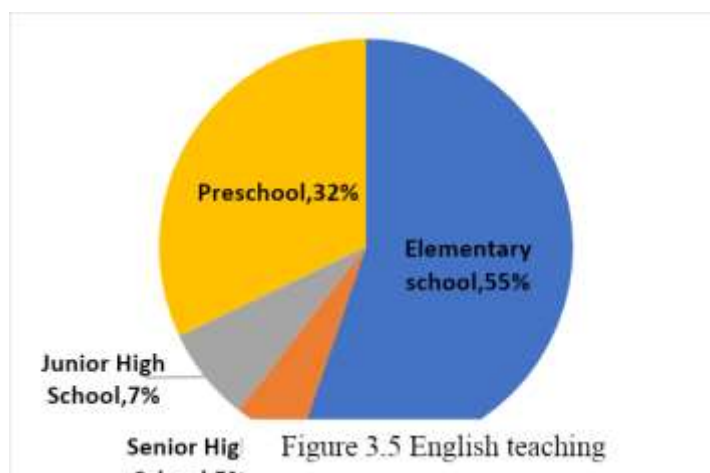
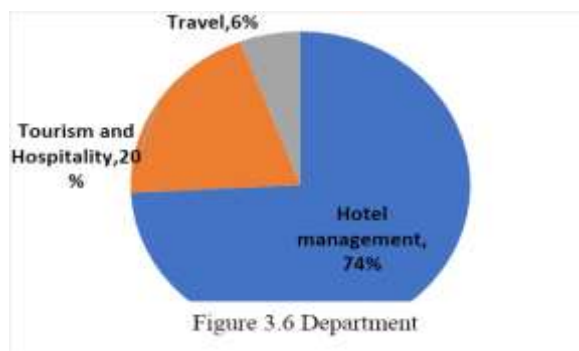


Figure 3.5 English teaching

The findings revealed that as much as 55% of the student-participants learned English from elementary school, and 32% learned from preschool. A smaller percentage was for junior high school (8 %) and senior high school (5 %). Despite the fact that English teaching is obligatory in secondary school, as much as 5% stated that they started learning English from senior high school because of the lack of English teachers in their junior high school. The lack of subject-specific teachers, such as English, on remote islands outside Java is evident. This phenomenon is also evident in elementary school. In primary education, English is taught as an extracurricular subject. However, the low competence and lack of teaching competencies of primary English teachers have been cited as severe obstacles. The study suggests that elementary English teachers and teacher educators require targeted preparation and additional training (Zein, 2014).

3.6.3.1.5 Department

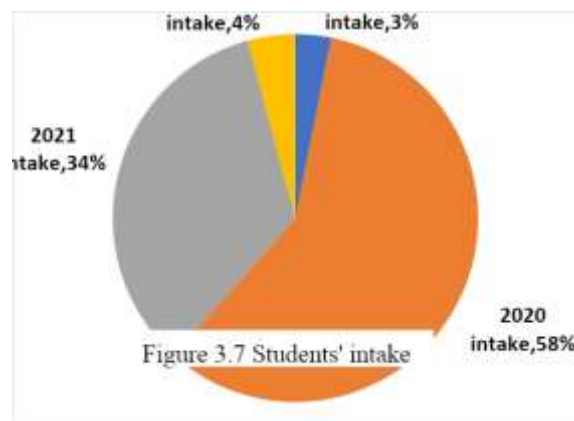
The next figure reports respondents' affiliations. As much as 74% of the respondents were from the hotel department, 20% were from the tourism and hospitality departments, and the rest were from the travel departments. The number of respondents was not equal because each department had a varied number of students, with hotel departments as the majority, followed by tourism and hospitality departments, and travel departments.



Additionally, several students' intakes were not in campus because they were still in the national and overseas industries for internship or on the job training (OJT). Nevertheless, the total number of respondents met the minimum sample of the study (Hair et al., 2018).

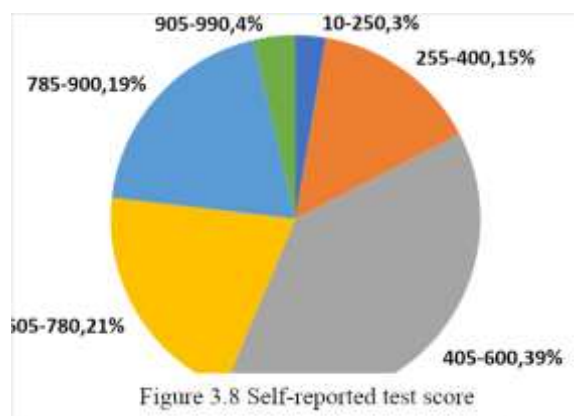
3.6.3.1.6 Intake

This figure elucidates the respondents' backgrounds. The majority of the student-participants came from the 2020 intake with 59%, which means that they were in the sixth semester. Additionally, 34% was from the 2021 academic year, and the rest was from the 2019 and 2022 academic years. Students from the 2019 academic year were in their final year and were likely to be in their second internship.



3.6.3.1.7 Self-reported Test Score

This item asked about the respondent's last test score, which were used as students' background ability. As mentioned in the previous section, students attended several prediction tests to observe their progress before they were ready to attend the TOEIC international test in the 7th semester.



To provide a general description of the respondents' performance, they mentioned self-reported score gains. As seen from the figure, a score range was given to see the students' level. A can-do list table was provided to interpret the scores.

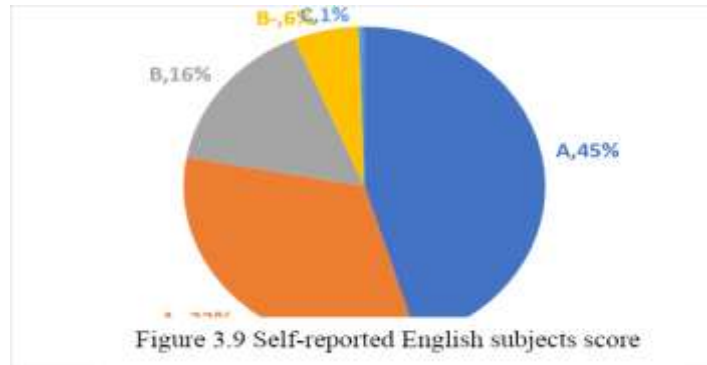
Table 3.9 Can-do-list Table

Score range	Proficiency level	Percentage
10-250	Novice	3%
255-400	Elementary	15%
405-600	Intermediate	39%
605-780	Basic working proficiency	20%
785-900	Advanced working proficiency	19%
905-990	Professional working proficiency	4%

The table shows that the majority sit at the intermediate level (39 %). At this level, the *can-do list* states that they can initiate and maintain face-to-face conversations and satisfy limited social demands (Rogers, 2003). Additionally, as much as 20% belonged to basic working and 19% had advanced working proficiency. At these levels, they are classified as having working proficiency, in which they can satisfy limited work requirements and routine social demands (Rogers, 2003). Lastly, the top achiever has a professional working proficiency of 4%, and Roger states that this category has advanced proficiency and is able to satisfy most work requirements with acceptable and effective language use. Despite the high achievers, as much as 18% of the respondents stated that they had elementary and novice proficiency. At these levels, they are claimed to satisfy the basic survival requirements (Educational Testing Service, 2015; Rogers, 2003).

3.6.3.1.8 Self-reported English Subject Score

Figure 3.9 reports the self-reported scores of the English subjects. The majority of the respondents (45%) reported that they got A, 33% got A-, and 22% got B and B-. These reports were almost in line with their self-reported test scores on the test preparation in **Table 3.9**.



3.6.3.2 Participants for Interviews

From the survey participants, four students responded and agreed to participate. The students involved in this case study came from various regions. Three of these students were from provincial regions, while the other participant was from city center. The participants were in their second year of their study. Referring to their teachers' categorization, the study included one advanced student, two intermediate proficiency students, and one elementary student. Students also provided details of their English learning experience as evidenced from the previous English course participation. In general, the students' reported test scores were in line with their teachers' assessments, as evidenced from the self-reported proficiency, and last mock-up test score range, which were taken in the first and second semester.

Table 3.10 Participants' profiles for Interviews (N=4)

Pseudo-nym	Gender	Academic year	High School location	Proficiency judged by teachers	Self-reported proficiency	Last mock-up score	TOEIC score target
Putri	Female	2 nd year	Urban	Intermediate	Intermediate	450	700
Vita	Female	2 nd year	Rural	Advanced	Intermediate	600	900
Senja	Female	2 nd year	Rural	Elementary	Elementary	355	600
Agi	Male	2 nd year	Rural	Intermediate	Intermediate	525	900

In particular, Vita, who were considered top students, achieved the highest scores. The students were also asked to self-assess their language skills. Interestingly, the top student tended to underestimate her own abilities, rating as having intermediate

level. The other participants' self-rated proficiency was similar with teacher's rate, while the low achiever admitted her novice proficiency, and she happened to come from rural area belong to 3T regions. These areas are named as 3T (*terdepan, terluar, tertinggal*) or further clarified as the frontier, outermost, and least developed districts.

After confirming their willingness to participate in the research, four students were chosen for interviews to investigate the third research question concerning the washback effect on learning. Their identities were anonymized and assigned as Putri, Vita, Senja, and Agi. Their backgrounds are elaborated below.

3.6.3.2.1 Case 1. Putri

Putri lived 15 kilometres from the university in South Jakarta. As the eldest of two siblings in a financially secure family, she was determined to learn English. Despite a government-mandated TOEIC test during vocational school, she enrolled in an English course before university. The course, provided by PT. ITC, included online learning about the purpose, format and strategies of the TOEIC. Putri's first exposure to English was in Year 1, which was characterised by fun games to build vocabulary. Although she found English both easy and difficult due to the complexity of the grammar, she valued it for future reading and writing. However, distractions, particularly her mobile phone, hindered her concentration and she turned it off to maximise her learning.

3.6.3.2.2 Case 2. Vita

Vita, a Bachelor of Tourism student, lived 15 kilometres from her city college and faced financial constraints in a family of three. Originally from Wonogiri, she attended a government high school 600 kilometres away. Financial constraints prevented her from taking private English classes during high school. Vita, who currently lives in South Jakarta, paid for her school fees through part-time work, supplementing her scholarship from the Foundation. Introduced to English in primary school, she initially struggled with speaking but found the subject generally

easy. She was fascinated by grammar, but admitted to having difficulty maintaining motivation for vocabulary. Despite financial challenges, Vita actively engaged in both tested (grammar) and non-tested (vocabulary) English skills.

3.6.3.2.3 Case 3. Senja

Senja comes from the Mentawai region, which is classified as a 3T area due to its frontier, outermost and least developed status. As the eldest of two siblings, her parents, farmers, face financial challenges. She grew up in Kep. Mentawai district in Siberut, South Mentawai, Senja's limited financial resources prevented her from studying English before university. The Ministry of Education and Culture's KIP Tuition Program is supporting her education in Jakarta, where she also works part-time to cover her living expenses. Senja started studying English in junior high school and struggled to grasp the material at first. While she finds English easy in general, she has difficulty understanding tenses, especially the simple past tense. Senja's commitment to overcoming challenges through diligent learning exemplifies the perseverance required in language learning and offers insights into the wider student experience of language learning in formal education.

3.6.3.2.4 Case 4. Agi

Agi, a confident and talkative student, eagerly volunteered for the research. Originally from Garut, he moved to Jakarta for his education, the fifth of six children in a financially stable family. Although he had never attended an English course during his school years, Agi's financial constraints led him to pursue self-directed learning. He currently lives in Jakarta and is responsible for supporting his retired parents and siblings. He attended SMAN 18 Garut for high school and opted out of English classes before college, relying on self-directed learning. Agi finds English easy and appreciates its accessibility through everyday life experiences. However, he dislikes the repetition of school lessons and is challenged by rigid learning structures that stifle creativity. To overcome this, Agi supplements his formal education with part-time jobs, English films, novels and books.

The participants were selected for several considerations. The first priority for selection of student participant was to ensure that they were affected by the English exit test, it means that at the 7th semester, they would have to attend the TOEIC international test as one of graduation requirements. The second consideration was from the participant' background. They represented varied level of region, and English proficiency. To assess their competency, several criteria were adopted. First, their previous prediction tests were used to map their initial data. Then, it was triangulated with the self-reported proficiency level, and teacher's judgement. Some of the students were from the researcher's classes, so the researcher could confidently assure on their varied socio-economic background and competency which can enhance the representativeness of the study. In general, the participants selection aimed to acquire cases that are rich in information and provide in-depth insights into the subject under study (Creswell, 2012;Freeman et al., 2007; Reybold et al., 2013).

3.7 Ethical Considerations

The research focused on a single study, and it is consistent with previous washback studies (Chen et al., 2020; Li et al., 2012). Ethical clearance was obtained from the Tourism VHE in Jakarta before the study began. The rector gave permission to conduct the research, and the letter of permission was distributed to stakeholders to obtain their consent to participate. Although some department heads and ESP teachers gave their consent, they were not included as participants because it was difficult to reach them amidst their busy schedules and heavy workloads. Similarly, the targeted student participants, particularly those from the Travel and Tours Department, who were from two academic year, were excluded as they were doing their industrial internships, both nationally and internationally. All participants who participated in the study gave informed consent. The use of pseudonyms ensured anonymity and confidentiality in the data presentation.

3.8 Data Collection Method and Instrumentation

This study requires the integration of both quantitative and qualitative methods in order to gain a comprehensive understanding of how the TOEIC test enacted as an English exit test affect various stakeholders. The strengths of each method are leveraged by combining findings from quantitative and qualitative research, resulting in a deeper understanding than if each approach was used in isolation (Schulze, 2003). In addition, combining qualitative and quantitative research allows the researcher to better understand the research problem (Creswell, 2014; Ivankova et al., 2006).

Table 3.11. Data Collection and Instrumentation

No	RQs	Participants	Data Collection	Instruments	Analysis
1	Policy rationales	4 participants (2 HoDs, 1 test coordinator, 1 language coordinator)	Semi-structured interviews	Interview list	Thematic analysis
2	Teaching washback	6 ESP teachers (3 of which were observed)	Document analysis Classroom observation Semi-structured interviews	ESP syllabus COLT Part A Interview list	Thematic analyses
3	Learning washback	190 students (4 of them were interviewed)	Survey Semi-structured interviews	Questionnaire: Part A. Demographic Part B. Washback Activity theory	Principle Component Analysis, Statistic Descriptive, Correlations Thematic analysis
Total		200 participants			

In the first research question, it seeks to explore the policymakers' reasons of adopting TOEIC as an English exit policy. To this end, semi-structured interviews were conducted with four policymakers. The second research questions dealt with the ESP teachers' view on the washback of TOEIC on their teaching practices.

To address the second research questions, there were three data supporting the findings: first, documents in the form of ESP syllabi, lesson plans, assessment, and teaching material were analysed to understand the comprehensive views on ESP teaching in the research settings. The competence criteria could not be compared with SKKNI competency requirement, since it was not available for the ESP subjects being discussed. To the researcher's views, there was only SKKNI standard for English for housekeeping. Second, classroom observations were conducted to validate the data from the documents. Third, interviews with ESP teachers were conducted to confirm the results of the previous data. The third research question dealt with students' perspectives on the washback of TOEIC on their learning. This problem was addressed with two data: questionnaire and interviews.

3.8.1 Interview

In this research, interviews were conducted to unshed the policymakers', ESP teachers, and students' perspective. The interview questions are given in **Appendix 1, 2, and 3**. The use of interviews in this study allowed researchers to gain insights and 'depth' information on participants' perspectives, experiences, attitudes, and emotions about a subject (Denscombe, 2007; Fraenkel et al., 2018). Additionally, interviews enable participants to offer historical information (Creswell, 2007). Heads of departments can discuss the background, benefits, and challenges of the program policy in the context of this study. To add the richness of the data, follow-up questions or 'probing techniques' was employed (Marshall & Rossman, 2016).

The instrument development for qualitative data collection takes a long process. Some considerations and steps for writing the interview questions are taken (Foddy, 1993; Gillham, 2000; Wilkinson & Birmingham, 2003). The list of interview questions is developed to answer the research questions, and the constructs are adapted from several references (Nguyen & Gu, 2020; Nguyen, 2015). The in-depth, recorded, semi-structured interviews, last approximately thirty to sixty minutes, were used to gather the views of all stakeholders.

To facilitate understanding and convenience, the interviews were mostly conducted in Bahasa. Thus, the interviewee could speak in whatever language they are most confident and comfortable (Murray & Wynne, 2001). This may also aid in the avoidance of misinterpretation of interview questions and responses. In short, it allows interviewees to express their ideas and experiences (Brinkmann & Kvale, 2018; Rowley, 2012), resulting in more authentic and more emotionally rich answers (Baumgartner, 2012; Welch & Piekkari, 2006).

To achieve data saturation, probing is used to seek for additional information (Robinson, 2023). Since the researcher taught in the research site, she had long been acquainted with most participants, which enabled her to build rapport (McGrath et al., 2019). The teachers' interview was conducted twice in this study, with the first and second taking place at different times to promote cross-checking and consistency. Furthermore, the second round of interviews are vital for assessing the reliability of the verbalization (Emilia, 2005). Member checking was performed to enhance credibility of the data (Birt et al., 2016) as well as promoting the participants' reflection (Candela, 2019).

3.8.2 Document Analysis

Document analysis was carried out during the initial phase of teachers' study to cater the second research question. The documents are crucial since they provide background information and a variety of facts, making them valuable for situating a study within its subject or field (Bowen, 2009), as well as serving as data triangulation (Merriam & Tisdell, 2016). Additionally, they are stable and 'non-reactive' data sources, allowing multiple readings without being influenced by researchers or research processes. Bowen (2009) further recommended that the analysis process should include the categorization of information into themes, similar to the examination of focus group or interview transcripts. To this end, the document was categorized based on five components covering (1) ESP curriculum, (2) course syllabus, and lesson plan, and (3) course material, and (4) English assessment. These components were scrutinized to investigate whether the ESP

classrooms followed the principles of communicative language teaching, focusing the development of communicative competence and real-life language use. The sample of the document can be seen in **Appendix 3**.

3.8.3 Questionnaire

Questionnaire was developed to unshed students' perspectives on the washback of TOEIC on their learning as stated in research question three. Questionnaire was adopted since it offers various advantages and, it is cost-effective and efficient, allowing for data collection in a single administration (Cohen et al., 2018; Dunne et al., 2005). Since the study adopts explanatory sequential analysis, the results of the quantitative data analysis informed the qualitative data collection.

Questionnaire development was performed in several steps. First, question writing, informal piloting and draft refinement, translation and expert judgement, piloting and draft refinement, and internal reliability and validity checking (See **Appendix 4**). To capture students' perspectives, the four-point Likert Scale was adopted; point 1 (one) showing strongly disagree and point four (4) strongly agree (Cohen et al., 2018). The absence of midpoint aimed to improve clarity (Chyung et al., 2017). Questionnaire comprised of three parts, with a total of 68 items. First, it asked students' background. Second, it asked how the TOEIC test affected their learning in terms of content of learning, total time of learning, learning strategies, learning motivation, test anxiety, learning outcome, and washback on online English test. Third, it asked several factors associated with learning, such as perception of the TOEIC test, beliefs about language learning, test pressure, and context factors. See **Appendix 4** for more details.

3.8.4 Classroom Observation (Onsite Mode)

The purpose of the observation was to collect data on classroom teaching practice in face-to-face or on-site mode. These observations were critical in washback studies since they provided a better insight of the content and character

of the programs, as well as their impact on learning, which questionnaires cannot fully capture (Alderson & Wall, 1993; Bailey, 1996). Additionally, observations allow for an assessment of the level of washback and show actions of students and teachers that may not be revealed in interviews or questionnaires (Hawkey, 2006; Saville, 2009). Eventually, it aids to reveal discrepancies between what teachers claim they do and what they actually do in the classroom (Al-Nouh, 2008; Orafi & Borg, 2009), thus presenting a more accurate view of reality (Cohen et al., 2018).

For this purpose, three classes taught by three different teachers were observed for four sessions each, making a total of twelve sessions. See **Table 4.5** for details. The researcher served as a participant-observer, observing and documenting visual and auditory perceptions (McMillan & Schumacher, 2001). As the sole observer, the researcher utilized data recording, both video and audio to capture the interaction the class, while at the same time she also took notes during the teaching and learning process to avoid any oversight. To avoid personal bias and personal judgement, she consulted the field note and recording with the teacher after the class was over. This step was done consistently to avoid missing any key events and to maintain data reliability.

In order to initiate a systematic and focused data collection process (Merriam & Tisdell, 2016), the researcher employed an observation sheet adapted from the Communicative Orientation of Language Teaching (COLT) (Allen et al., 1983; Read & Hayes, 2003). The COLT instrument consists of two parts: Part A focuses on the behaviour and interactions of teachers and students in the classroom, while Part B describes the verbal exchanges between teachers and students during specific activities. This study focused specifically on Part A of the COLT observation scheme, with the aim of establishing a clear link between measured teaching practices and variations in educational outcomes.

3.9 Data Analysis

This study adopted a variety of data collection; thus, a combination of qualitative and quantitative data analysis was used in this research. First the quantitative data

resulted from the questionnaire were analysed with univariate and multivariate data analyses with SPSS 20 and JASP for Principal Component Analysis, statistic descriptive, and correlational among the variables under study. The descriptive statistics were presented based on the factors found through PCA, and they were presented for percentage and mean in **Sub-chapter 4.3.1**

The qualitative analysis, following the thematic analysis (Miles et al., 2014), involved a thorough review and rereading of the transcripts of the interview, field note, and documents. Then, relevant excerpts corresponding to central themes were identified, coded and notable points documented. Similar codes were grouped together to facilitate interpretation. The process involved transcription, categorization and interpretation, in line with the approach suggested by Emilia (2005) by comparing data from one category to the next in order to obtain condensed information (Emilia, 2005). The accuracy of the transcribed data was checked before presenting extracts in a non-verbatim format. In addition, peer debriefing was incorporated to enhance credibility by seeking feedback from an experienced colleague to ensure alignment between findings and participant data. The findings are supported by extracts from the transcripts, presented in a concise format. This method involves omitting unnecessary words and retaining only the gist of participants' quotes. The complete data analyses of the three research questions were presented in **Figure 3.10** below.

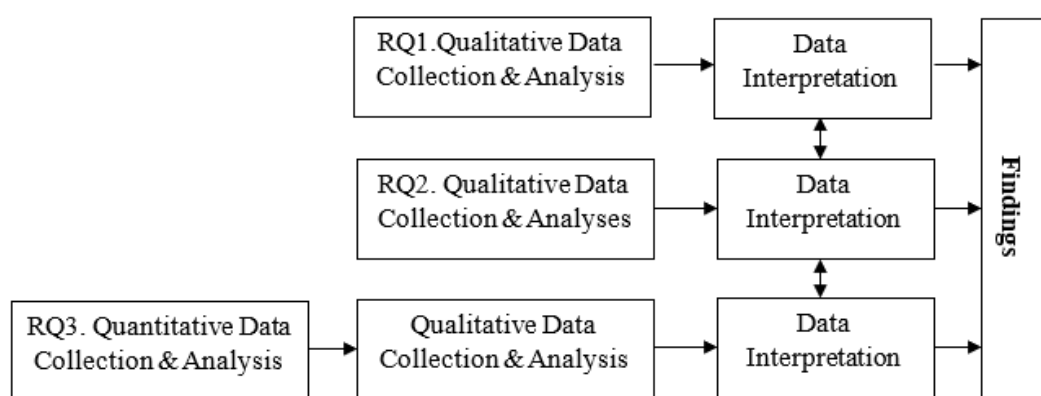


Figure 3.10. Data Analyses

3.10 Examining washback from Activity theory (AT)

As informed in **Sub-chapter 2.11**, this study also seeks to analyse the washback of TOEIC using activity theory. The selection of AT as the conceptual framework in this study lies on its ability to portray the conceptualization of individuals and their environment as a holistic unit of analysis (Yamagata-Lynch & Haudenschild, 2009). Additionally, the multidisciplinary interest in cultural practices and practice-bound cognition can benefit greatly from Activity Theory (Engestrom, 2000). Initially, the coding process for each transcript was then guided by a top-down coding technique based on AT theoretical components (Engeström, 2015). The identified nodes are presented in the table below.

Table 3.11 Statements node categories (Adapted from Engestrom, 2000)

Category	Explanation	Component
Object and outcome	the ultimate objective or result that the subjects are working toward, outcome is the expected achievement.	Object: Students' English skills; Outcome: Students' enhanced English competency to match industry demand
Subject	Those who participate in an activity and are essential to how it is understood.	Students' preparing TOEIC in VHE context
Division of labour	the shared obligations of community members, including roles and relations of power.	How students learn English
Rules	both explicit and implicit standards and regulations that may either aid or hinder the subjects in accomplishing their goals.	Exit test policy, TOEIC benchmark, campus policy

Tools	The methods or resources that the subjects utilize to accomplish their objectives.	Learning resources, instruction, curriculum, language learning and teaching
Community	institution or group that the subjects are a part of while participating in the activity	Students, teachers, HoDs, a language coordinator, a test coordinator, family

The researcher operationalized the AT theory with the following steps (Mwanza, 2001).

Stage 1 – Model the situation being examined ?

Stage 2 – Produce an Activity System of the situation ?

Stage 3 – Deconstruct the situation’s Activity System ?

Stage 4 – Generate Research Questions* ?

Stage 5 – Conduct a detailed investigation ? Stage 6 – Interpret Findings

Activity Theory has been used in research on washback by examining the processes and products of human activity related to high-stakes testing, and analyzing the goal-directed actions and operations that are essentially motivated by need. The theory is used as a framework for investigating the relationship between learner behavior and the impact of testing on learning. In this research, the focus is not only on the test itself, but also on the wider political, cultural, and social contexts of testing and their influence on student learning. Activity Theory allows researchers to explore how situated learner goals and actions may contribute to the effects that tests have on learning. By analyzing the interactions between learners and their communities, as well as the motivations and needs behind their actions, researchers can gain a deeper understanding of the impact of testing and the processes that lead to that impact. In terms of instrument conceptualization, Activity Theory can be used to develop and evaluate assessments and other instruments related to language learning and testing. By focusing on the context of language use and the interactions between learners and their communities, assessments and

instruments can be designed to better reflect the real-world situations in which language is used and acquired.

In this study, each component making the AT triangle is specified, and elaborated. Then, the analysis looked at how each component interact with others to explain the complex nature of washback on students' learning process. Two student cases were analysed based on this theory: Vita and Agi.

3.11 Validity and Reliability

Data quality control is performed by examining the validity and reliability of the instruments. Validity aims to investigate if an instrument actually provides a measure of what it purports to measure (Taherdoost, 2016). To this end, validity is determined from the way a scale developed, its capacity to predict certain events, or its relationship to measures of other constructs, and there are three fundamental categories of validity: content, criterion, and construct validity (DeVellis & Thorpe, 2022). The following paragraphs elaborates the three types of validity as stated previously.

3.11.1 Content Validity

Expert judgement was used to assess content validity (Olson, 2010). The questionnaire draft was validated for both language and content by the research supervisors. The English version of the questionnaire was then translated into Bahasa to facilitate students' understanding. The translated questionnaire is attached in **Appendix 6**.

3.11.2 Construct-related Validity

To investigate the construct validity of the questionnaire, a Principal Component Analysis (PCA) with *Jeffrey's Amazing Statistics Program* (JASP) and SPSS 20, was conducted to address the construct of the questionnaire. PCA aims to substitute a group of interrelated observed variables with fewer uncorrelated latent variables, called "components," that contain most of the data in the original dataset (O'Sullivan et al., 2021). The primary objectives of principle component analysis

(PCA) are to (a) extract the essential information from the data table, (b) minimize the data set size by maintaining just the significant information, (c) simplify the data set description, and (d) investigate the structure of the observations and variables. In summary, PCA is preferred to Exploratory Factor Analysis (EFA) for construct-related validity because of its simplicity (Brown, 2009). This method of exploratory descriptive analysis involves summarizing and aggregating related variables (Santos et al., 2019). PCA also has the ability to deal with multicollinearity (Forsberg & Rantala, 2020), and data reduction (Brown, 2009a). These features make it a valuable tool for assessing the underlying structure and validity of measures.

This analysis explains how well the items in questionnaire capture the substance of the constructs to be investigate (Saunders et al., 2009). The steps and statistical process can be seen in **Appendix 8**. One factor, learning from friends, had a Cronbach's Alpha (CA) .584, thus it was not reliable, thus it was not included in the discussion. The result of PCA is seen in Table 3.12

Table 3.12 Composite Variable

No	Question	Original items	Component	Factors found	Item making up the factors	CA
A	Impact on learning					
1	To what extent has the TOEIC changed your content of learning?	1, 2, 3, 4, 5	PC10 (Direct test prep strategies)	Taking more test-taking strategies & MCQs	C3, C4	.631
			PC11 (Content)	Content reduction	C1, C2	.675

2	To what extent has the TOEIC test changed your total-self-study time on the following skills?	6,7,8,9,10,11,12	PC4 (Time for language skills)	Time devoted to tested skill	TLS1, TLS2, TLS5, TLS6	.753
				Time devoted to non-tested skill	TLS7, TLS3, TLS4	.702
3	To what extent has the TOEIC test changed the amount of time for learning activities?	13,14,15,16,17,18,19	PC6 (Time for learning activities)	Time for form-focused activities	TLA5, TLA6, TLA7	.766
				Time for meaning focused activities	TLA3, TLA2, TLA1, TLA4	.756
4&1 2	To what extent does the TOEIC test affect your learning strategies?	20,21,22,23,24,25	PC3 (context and learning strategies)	Learning strategies from teachers	LS5, CF3, TP4	.654
				Test preparation and learning strategies	LS3, LS1	.606
				Learning strategies from friends	LS6, CF2	.584
5	To what extent does the TOEIC test affect English learning motivation?	26,27,28,29,30,31,32,33,34,35	PC1 (Motivation)	Intrinsic learning motivation	LM2, LM10, LM9, LM7, LM8,	0,816

				Extrinsic learning motivation	LM7, LM6, LM5, LM1, LM3	.873
6 & 11	To what extent does the TOEIC test affect your feeling?	36,37,38, 39	PC5 (Test anxiety and pressure)	Test anxiety and pressure	TA2, TA1, TA4, TP1	.726
				Academic and graduation concerns	TA3, TP2	.666
7	To what extent does the TOEIC test affect your achievement?	40,41,42, 43,44,45	PC2 (Learning outcome)	English skill improvement	LO2, LO5, LO3, LO4	.895
				Overall English competency improvement	LO1	
				Test performance improvement	LO6	
B	Factors related to learning					
8	To what extent has the TOEIC	46,47,48, 49,50	PC12 (Online test problems)	Unfamiliarity with test interface	OTW1	

	online test affected you?			Focus problem during online test	OTW2	.654
			PC9 (online infrastructure)	Incompatible devices	OTW5	.763
				Unstable connection	OTW4	
				Limited internet access	OTW3	
9	What do you think about the TOEIC as one of graduation requirement?	51,52,53,54	PC8 (Test perception)	TOEIC to measure English knowledge and communication skills	TP1, TP2	.873
				TOEIC as learning goal & employment	TP3	
				TOEIC score as an indicator of English competency	TP4	
10	What are your beliefs about teaching and learning English?	55,56,57,58	PC7 (Learning beliefs)	Belief that English learning improves test score	LB3	.756
				Belief in the need of	LB4	

				learning language skills	
				Belief that language learning is for communication	LB1

The results of PCA were then labelled based on the item making up the factors. The findings revealed that as many as 59 items have made up 12 components, and 7 other items did not cluster any components. These items were then, dropped, and not included in the data analysis and findings. Having been labelled, the components were then mapped to address the 12 questions in the questionnaire. It is worth noting that some components address several questions, such as component 3 which deals with context and learning strategies (Question 4 and 12), and component 5 which revealed the test anxiety and pressure (Question 6 and 11). To this end, the twelve-questions have now adjusted into 10 questions as seen in **Table 3.12**.

3.11.3 Internal Reliability

Two data analysis steps are performed to ensure the quality of the data collected and help explain the research findings. First, to investigate the internal consistency of the questionnaire, the reliability test of the questionnaire data is calculated, with Cronbach's for all items with four-point Likert-scale items. In addition to item validity, reliability checking is conducted by looking at the Cronbach's Alpha (CA). Reliability is seen as the consistency of study outcomes across time, instruments, and respondent groups (Cohen et al., 2018) or called item consistency for short (Ary et al., 2010).

Additionally, the reliability analysis with Cronbach's Alpha, mostly showed high reliability on most variables, and one variable showed low reliability (See **Appendix 7**). This internal consistency measurement was established by Lee

Cronbach in 1951, with a range of 0 -1 (Cronbach, 1951). The Cronbach' s Alpha of .6 or higher is widely believed as reliable in social science studies (Shi et al., 2012). In this study, one construct, learning strategies from friends, found in LS6 and CF2, had a CA of lower than 0.6 as the cut-off point. Thus, removed from the analysis since the items were not reliable. Other construct CA ranged from .606 to .895 which were classified as having a good reliability.

3.12 Trustworthiness and Transferability

The nature of trustworthiness in qualitative research deals with validity concept, thus the study is regarded as “fit for purpose” (Hall, 2013). Trustworthiness guarantees the findings that can be relied on and suitable with the research questions, methods, strategies, data analysis, clarity and detail information of the research process, appropriate interpretations, and clear reporting procedure. Trustworthiness is maintained with interviews recording, transcribed verbatim, and given to the interviewees for member checking. Additionally, to aid understanding, the interviews are conducted in Bahasa Indonesia to ensure successful communication. Triangulation is also conducted by checking the information from different sources. To this end, interviews from 4 alumni both working for foreign and domestic industry, employers (2 hotel general managers, and 1 HRD manager) and survey on trainee language competence from 105 internship providers (foreign and domestic) were provided in Chapter 4. It is conducted by adopting multiple perceptions to clarify meanings. To enhance the validity and understanding, a data source triangulation was done by contrasting and comparing data gathered from various sources, including students' journals, classroom observations, text analyses, and interviews (Mertens & Hesse-Biber, 2012; Youngs & Piggot-Irvine, 2012).

As with transferability, findings in qualitative research are not for generalization of results, it aims to extend understanding of the quantitative gain. In this case, this research is appropriate for research in similar setting that is stakeholders in VHE in travel and hotel departments.

3.13 Concluding Remark

Chapter III lays out the research approach employed in the current study. The chapter opens with a description of the research design, emphasizing the use of both quantitative and qualitative approaches. The population and sample are thoroughly discussed, with a particular emphasis on the selection criteria. The research technique was rigorously outlined, including the quantitative phase, which included questionnaires for ESP students, validity and reliability measurements, as well as data collection and analysis procedures. Furthermore, the qualitative portion of the study describes the inspection of documents, classroom observations, and interviews with various parties, including the Head of Department, the test coordinator, the language coordinator, ESP teachers, and students. The chapter also describes how Activity Theory was used to study washback, emphasizing the importance of trustworthiness and transferability considerations. Ethical considerations and research authorization are addressed to maintain research integrity. To sum up, this chapter provides a solid foundation for the following chapters by presenting a comprehensive description of the methodology adopted.