

FACE VALIDITY OF THE INSTITUTIONAL ENGLISH BASED ON THE COMMON EUROPEAN FRAMEWORK OF REFERENCE AT A PUBLIC UNIVERSITY IN VIETNAM

Nong Thi Hien Huong*

*Thai Nguyen University of Agriculture and Forestry
Tan Thinh, Thai Nguyen, Vietnam*

Received 17 September 2019

Revised 23 December 2019; Accepted 14 February 2020

Abstract: In language testing and assessment, face validity of a test is used by learners and is probably considered as the most commonly discussed type of test validity because it is primarily dealt with the question of whether a test measures what it is said to measure. Therefore, this study investigates students' and English lecturers' perceptions toward the Institutional English Test based on the Common European Framework of Reference administered in a public university in Vietnam. A survey of 103 students and 20 English lecturers from the Institutional Program was conducted. A questionnaire with 7 main concerns – weightage, time allocation, language skills, topics, question items, instructions and mark allocations was used to collect data. All responses were analyzed through descriptive statistics. The results showed that the Institutional English Test based on the Common European Framework of Reference had satisfactory face validity from both the students' and lecturers' opinions; consequently, the Institutional English Test is perceived as a good test to measure students' English abilities.

Key words: language testing, test validity, face validity, test validation

1. Introduction

In our globalized world, being able to speak one or more foreign languages is a prerequisite, as employers on a national as well as on an international scale pay attention to the foreign language skills of their future employees (Kluitmann, 2008), focusing mostly on English.

Therefore, English nowadays has been gaining an important position in many countries all over the world. English is not only a means but also an important key to gain access to the latest scientific and technological achievements for developing countries such

as Vietnam, Laos, Cambodia and Thailand. Furthermore, it is estimated that the number of native English speakers is approximately 400 million to 500 million; more than one billion people are believed to speak some forms of English.

Campbell (1996) claimed that although the numbers vary, it is widely accepted that, hundreds of millions of people around the world speak English, whether as a native, second or foreign language. English, in some forms, has become the native or unofficial language of a majority of the countries around the world today including India, Singapore, Malaysia and Vietnam.

In Vietnam, the Vietnamese government has identified the urgent socio-political,

* Tel.: 84-984 888 345

Email: nongthihienhuong@tuaf.edu.vn

commercial and educational need for Vietnamese people to be able to better communicate in English. In line with this aspiration, all Vietnamese tertiary institutions have accepted English as a compulsory subject as well as medium of instruction for academic purposes. This development has given rise to the need to teach and measure students' command of English at institutional level. However, the issue that is often raised in relation to in-house language test is validation because the locally designed language tests are disrupted by the fact that they do not indicate the features of language skills tested and hardly tap the students' language abilities (Torrance, Thomas, & Robison, 2000).

According to Weir (2005), test validation is the "process of generating evidence to support the well-foundedness of inferences concerning trait from test scores, i.e., essentially, testing should be concerned with evidence-based validity. Test developers need to provide a clear argument for a test's validity in measuring a particular trait with credible evidence to support the plausibility of this interpretative argument" (p. 2). Therefore, test validation has been considered as the most important role in test development and use and should be always examined (Bachman & Palmer, 1996). Face validity is one of the components in test validation and is probably the most commonly discussed type of validity because it was primarily dealt with the question of whether a test looked as if it measured what it was said to measure (Hughes, 1989).

Bearing this in mind, this study aims to investigate the face validity of the Institutional English Test (IET) based on the Common European Framework of Reference at a public university in Vietnam. Most of the previous studies in accordance with language test validation have been derived from the views of educators or researchers; however, in this study the perceptions of both students

and English language lecturers as important groups of stakeholders were collected (Jaturapitakkul, 2013; Kuntasal, 2001; Samad, Rahman, & Yahya, 2008). The results might shed some lights on English language testing and could primarily inform ways to improve current in-house English language test.

2. Literature review

2.1. *The importance of language testing*

Language testing and assessment is a field under the broad concepts of applied linguistics. This field has been rooted in applied linguistics because it is related to English language learners, test takers, test developers, teachers, administrators, researchers who have great influences on teaching and learning English in the world (Bachman, 1990). He explains in detail that testing is considered as a teacher's effective tool contributing to the success of teaching English in the classroom as well as helps him or her produce the exact and fair evaluation of students' ability and the performance of the language (Bachman, 1990).

Sharing the same view, McNamara (2000) defines language testing as an aspect of learning that helps learners to grasp the knowledge that they have missed previously and the teacher to understand what can be done in subsequent lessons to improve teaching. To (2000) presents language testing as a useful measurement tool which test validation can assist in creating positive wash back for learning through providing the students with the feeling of competition as well as a sense that the teachers' assessment coincides with what has been taught to them.

In the same token, Davies (1978) emphasizes that "qualified English language tests can help students learn the language by asking them to study hard, emphasizing

course objectives, and showing them where they need to improve” (p.5). Similarly, McNamara (2000) highlights some important roles of language testing which have been applied popularly in educational system and in other related fields to assist in pinpointing the strength and weakness in academic development, to reflect the students’ true abilities as well as to place the student in a suitable course.

Additionally, language testing helps to determine a student’s knowledge and skills in the language and to discriminate that student’s language proficiency from other students (Fulcher, 1997). In the same vein, Hughes (1989) also states that language testing plays a very crucial role in the teaching and learning process because it is the final step in educational progress. Thus, to use tests to measure the educational qualities, the administrators should build important and qualified testing strategies which assist evaluating learners’ performance, teaching methods, materials and other conditions in order to set up educational training objectives (McNamara, 2000).

In short, language testing has assumed a prominent measurement in recent effort to improve the quality of education because testing sets meaningful standards to schooling systems, teachers, students, administrators and researchers with different purposes. Furthermore, language testing has enriched the learning and teaching process by pinpointing strengths and weaknesses in the curriculum, program appropriations, students’ promotion as well as teachers’ evaluation.

2.2. *Face validity*

Messick (1996, p.13) defines test validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationale support the adequacy

and appropriateness of inferences and actions based on test scores and other modes of assessment”. In other words, test validity or test validation means evaluating theoretically and empirically the use of a test in a specific setting such as university admission, course placement and class or group classification.

Bachman (1990) also emphasizes that overtime, the validity evidence of the test will continue gathering, either improving or contradicting previous findings. Henning (1987) adds that when investigating the test validity, it is crucial to validate the results of the test in the environment where they are used. In order to use the same test for different academic purposes, each usage should be validated independently.

Crocker and Algina (1986) highlight three kinds of test validity: Construct validity, Face validity and Criterion validity. In the early days of language testing, face validity was widely used by testers and was probably considered as the most commonly discussed type of test validity because it was primarily dealt with the question of whether a test measures what it is said to measure (Hughes, 1989). In a common definition, face validity is defined as “the test’s surface credibility or public acceptability” (Henning, 1987, p.89). In other words, face validation refers to the surface of a test such as behaviors, attitudes, skills, perceptions it is supposed to measure. For example, if a test intends to measure students’ speaking skills, it should measure all aspects of speaking such as vocabulary, pronunciation, intonation, word and sentence stresses, but if it does not check students’ pronunciation, it can be thought that this test lacks face validity.

Heaton (1988) states that the value of face validity has been in controversy for a long time and has considered as a kind of scientific conceptual research because this validation

mainly collects data from non-experts such as students, parents and stakeholders who give comments on the value of the test. In the same view, several experts who have emphasized the importance of face validity, state that this validity seems to be a reasonable way to gain more necessary information from a large population of people (Brown, 2000; Henning, 1987; Messick, 1994). More specifically, these researchers highlight that using face validity in the study encourages a large number of people to take part in a survey, so it can be easy to get valuable results quickly. Therefore, Messick (1994) concludes that face validity must be among the various validity aspects in language testing and test validation.

To sum up, face validity examines the appearance of test validity and is viewed as a quite important characteristic of a test in language testing and assessment because this evidence helps the researchers gain more necessary information from a large population as well as get quicker perceptions about the value of the test.

2.3. Theoretical framework

As far as concerned, validity has long been acknowledged as the most critical aspect of language testing. Test stakeholders (test takers, educators) and other test score users (university administrators, policy makers) always expect to be provided with the evidence of how test writers can determine and control criteria distinctions between proficiency tests applied with different levels. Therefore, there is a growing awareness among these stakeholders of the value of having not only a clear socio-cognitive theoretical model to support for the test but also a means of generating explicit evidence on how that model is used and taken in practice. The socio-cognitive framework for developing

and validating English language tests of Listening, Reading, Writing and Speaking in Weir's (2005) model of conceptualizing test validity seem to meet all the demands of the validity in the test that test stakeholders want to use in the public domain. Sharing the same view, O'Sullivan (2009) emphasizes that the most significant contribution to the practical application of validity theory in recent years has been Weir's (2005) socio-cognitive frameworks which have had influenced on test development and validation. Similarly, Abidin (2006) points out that Weir's (2005) framework combines all the important elements expected of a test that measures a particular construct in valid terms. Table 1 presents an outline of the socio-cognitive framework for validating language tests.

Weir (2005) proposed four frameworks to validate four English language skills: Listening, Reading, Writing and Speaking. In each framework, Weir (2005) put emphasis on validating test takers' characteristics, theory-based validity (or cognitive validity) and other types of validation. At the first stage of design and development of the test, test-taker characteristics, which represent for candidates in the test event, always focus on the individual language user and their mental processing abilities since the candidate directly impacts on the way he/she processes the test task. In other words, in this stage, the important characteristics which are related to the test-takers may have potential effect on test, thus the test-developers must consider the test-takers as the central to the validation process first. The view of test taker characteristics under the headings: Physical/ Physiological, Psychological, and Experiential was presented in details by Weir (2005) in Table 1.

Table 1. Test-taker characteristics framework suggested by Weir (2005)

Physical/ Physiological	Psychological	Experiential
- Short-term ailments: Toothache, cold...	Personality	- Education
	Memory	- Examination experience
-Long term illnesses: hearing age, sex, vision...	Cognitive style	- Communication experience
	Concentration	- Target language country residence
	Motivation	
	Emotional state	

Another important test validation component which is highly recommended by the researcher is theory-based validity or Cognitive validity (Khalifa & Weir, 2009). It focuses on the processes that test-takers use in responding to test items and tasks. It should be emphasized that face validity is a part of cognitive validity in test validation. This validity requires test -takers to find out if the internal mental processes that a test elicits from a candidate resemble the processes that he or she would employ in non-test conditions. Furthermore, cognitive includes executive resources and executive process. Executive resources consist of linguistic knowledge and content knowledge of the test-taker. The test-taker can use grammatical, discursal, functional and sociolinguistic knowledge of the language in the test. These resources are also equivalent to Bachman’s (1990) views of language components. Weir (2005) defines language ability as comprising of two components: language knowledge and strategic competence that will provide language users with the ability to complete the tasks in the test. He also emphasizes that there are two main methods to explore the cognitive validity. Firstly, cognitive validity can be checked through investigating test-takers’ behaviors by using various types of verbal reporting (e.g., introspective, immediate retrospective, and delayed retrospective) in order to stimulate their comments on what they often do in Listening, Reading, Writing and

Speaking tests (Huang, 2013; Shaw & Weir, 2007). Secondly, a test’s cognitive validity can be examined through learners’ perceptions on Listening, Reading, Writing and Speaking tasks in their real life situation (Field, 2011). It can be noted that the two methods in cognitive processing will be selected individually, but it is suggested from test developers’ perceptions that whether they want to select the first or the second method, the process of performance of the test should be more like the process in the real life. Therefore, it can be said that investigating face validity is as important as evaluating the content or predictive validity of an in-house language test. However, there have been still some limitations in previous studies in terms of content and methodology. For illustrations, several studies (Advi, 2003; Ayers, 1977; Dooley & Oliver, 2002; Huong, 2000; Mojtaba, 2009; Pishghadam & Khosropanah, 2011) paid much attention to investigate the content validity and predictive validity of an in-house test more than face validity. To be more specific, the researchers tended to measure test scores rather than other perceptions about knowledge, skills or other attributes of students. Messick (1995) emphasized that the meaning and values of test validation apply not just to interpretive and action inferences derived from test scores, but also inferences based on other means of observing. This means that investigation of face validity will create much more validity for the tests. For these reasons above, this

study attempts to fill the limitations stated above by employing the qualitative method to investigate the face validity of the IET at a public university in Vietnam in order to improve the quality of education; pinpoint strengths and weaknesses in the curriculum and test administrations.

2.4. Previous studies on face validity

Some previous studies in language testing have already been conducted in an attempt to analyze the different aspects of test validation. McNamara (2000) points out that insights from such analysis provide invaluable contribution to defining the validity of language tests. Exploring how other researchers have investigated the face validity of a language test can shed light on the process followed in this research.

To begin with, Kucuk (2007) examined the face validity of a test administered at Zonguldak Karaelmas University Preparatory School, in Turkey. 52 students and 29 English instructors participated in this study. The researchers used two questionnaires and students' test scores. The instructors and students were given questionnaires to ask for the representative of the course contents on the achievement tests. All data were analyzed through Pearson Product Moment Correlation and Multiple Regression. The results showed that even though it appeared that Listening was not represented on the test, both English instructors and students still agreed that the tests still possessed a high degree of face validity. The results showed that the tests administered at Zonguldak Karaelmas University Preparatory School, in Turkey were considered valid and the test scores could be employed to predict students' future achievement in their department English courses.

Another research on face validity goes for Lee and Greene (2007) who explored the

face validity of an English Second Language Placement Test (ESLPT) by using both qualitative and quantitative data. The study was conducted with the total of 100 students and 55 faculty members at University of Illinois at Urbana Champaign, in the United States. A self-assessment questionnaire was administered to elicit students' own assessments of their academic progress and performance at mid-semester. Furthermore, the faculty evaluation questionnaire was given to 55 staff members to get the opinions about students' English proficiency, academic performance in the course, and the extent to which students' level of proficiency caught up with their performance in the academic course. Interviews with 20 students and 10 faculty members during their office hours were conducted individually. The results showed the ESLPT did not correlate considerably with faculty members' ratings of performance in content courses ($r=.14$). The findings indicated that international graduate students' English difficulties had less effect on students' academic performance than was expected, because of such other factors as sufficient background knowledge and lecture type courses during their first-semester studies.

A study was conducted by Şeyma (2013) investigating how well various assessment practices (placement test, midterms, quizzes, and readers) of the preparatory year English program in the Department of Foreign Languages predict the success of students for TOEFL ITP at TOBB University of Economic and Technology (TOBB ETU). The researcher used a questionnaire to investigate both the instructors' and students' opinion on the effectiveness of these assessment practices on TOEFL ITP and the scores of 337 students to find out the relationship between in-house assessment practices and TOEFL ITP. All data was analyzed through Pearson Product Moment Correlation and Multiple Regression.

The result revealed that students believed that mid-term exams were the most effective and beneficial assessment practice which helps students get higher scores from TOEFL ITP. Whereas, lecturers believed that quizzes were more effective for students' success in TOEFL ITP test.

3. Research questions

The study aims to investigate the face validity of the IET based on the Common European Framework of Reference at a public university in Vietnam through both students' and English language lecturers. The study intends to answer the following research questions:

1. What are students' opinions about the face validity of the IET?
2. What are English language lecturers' opinions about the face validity of the IET?

4. Significance of the study

With the continuous use of a language test for its locally designed purposes, it is importantly noted that validity becomes a property of the test (Bachman, 1990; McNarama, 2000; Davies, 1989). Therefore, the results of the study can be hoped to contribute the following:

- This study is one of the few, which will shed light on the review of literature on language testing practices and provide educators with more information related to test validation.
- This present study may be valuable for other institutions in their endeavor to validate in-house tests, to justify the correctness of their interpretations. They may take this study as a guideline to examine the quality of their locally-designed assessment tools. Most

importantly, it will contribute useful insights to English language teaching and learning, especially in-house English test validation and prevent the mismatch between learners' true performance and their test scores.

- It helps test designers and educational decision makers to check to what extent the course content can be adequately represented in the test content by observing the distribution of the frequencies among the content areas for future exam construction.
- For Vietnam context, this study is undertaken with the hope of providing the test validation guideline for local university English language tests as well as improving undergraduate students intakes at local universities.
- For universities, this study may provide the validity evidence for the in-house language tests. If the IET is found to be valid, this could be the potential for other universities to venture into the test validation, encourage students to improve their English skills and competencies which are required to succeed in the respective program.

5. Methodology

5.1. General direction of methodology

The research question is checking the face validity of the IET through the students' and English lecturers' perceptions. This stage needs to take place after the students have just completed their IET and the lecturers have just finished teaching their third -semester English course. During the first stage, both students and English lecturers would be required to assess the IET components: Listening, Reading, Writing, and Speaking, assess IET format and weighting and then respond to the data collection instruments.

In fulfilling the requirements for carrying

out this study, the research figured out the general direction of Methodology that the study would undertake in Figure 1 below:

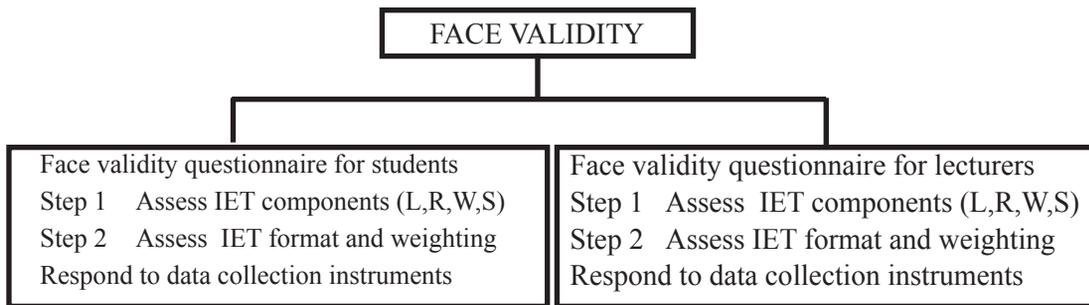


Figure 1. General direction of methodology

5.2. Participants

The participants of the main study consisted of 103 students who had completed their English course. The participants' ages ranged from 18 to 22 years. Furthermore, 20 English lecturers participated in the survey for face validity investigation. These English lecturers were teaching English at a public university in Vietnam and their ages ranged from 30 to 50 years. More importantly, they all have had experiences in teaching, designing the English tests as well as assessing the students' language ability.

5.3. The IET face validity questionnaire

Questionnaires have been the most frequently used data collection method in educational evaluation research because they help to gather information on knowledge, attitudes, opinions, behaviors and other information from a large number of people in a short period of time as well as at a relatively low cost (McLeod, 2014). Bearing this in mind, the questionnaire is used to collect the students' and lecturers' opinions about the IET in order to investigate the face validity of the IET as well as to answer the research questions. Some face validity questionnaires (FVQ) from previous studies (To, 2001; Jaturapitakkul, 2013; Kucuk, 2007; Kuntasal, 2001; Kuroki, 1994; Wang, 2006) were

collected. The focus on test weightage, time allocation, the representation of language skills, the representation of topics, the clarity of questions, the clarity of instruction and mark allocations in these previous FVQ was listed in order to gather necessary items which are useful for examining the opinions about the validity of a language test. Next, the first draft of the questionnaire for face validity of the IET was produced from these previous studies and then refined to make sure that the adaptation of the instrument would meet the requirements of investigating the lecturers' and students' opinions about the validity of the IET.

The face validity questionnaire of the IET is drafted for two groups of the participants in this study: Students and English lecturers. It consists of two main parts: Cover letter and Content of the questionnaire.

Cover letter

The construction of the consent cover letter aimed to gain permission to conduct the data from the students and the lecturers. The students' FVQ is the same as in the lecturers' FVQ.

The consent cover letter is the first part of the instrument construction. It begins with a brief introductory statement about the study and the researcher. Furthermore, the promise of

confidentiality is compiled in this letter to help the participants understand that their responses will not be in any case that affects their academic study or their academic career. Finally, contact and return information that is helpful to deal with queries during the data collection procedure is also included in the letter

Questionnaire content

Questionnaire content is the main part of the instrument construction. It consists of two sub-sections: Background information and Test components.

Section A is the first section which aims to ask for the students' and lecturers' background information. For the students, 9 questions were designed to ask for their full names, matrix number, gender, age, email, cell phone number, years of learning English and English speaking countries residence. For the lecturers,

7 questions related to background information were created to explore their full name, gender, age, email, cell phone number, educational qualifications and years of teaching English at a public university in Vietnam.

Section B is the most important section which aims to gather information on the IET components which are comprised of Listening Test, Reading Test, Writing Test and Speaking Test. This section contains 28 questions. Each component contains 7 questions which ask for the opinions on the weightage, time allocation, the representation of skills, the representation of topics, the clarity of questions, the clarity of instructions and the mark allocation. The responses to the questions are ranked from Strongly disagree to Strongly agree.

The framework of the adapted instrument for face validity of the IET is presented in Figure 2 below:

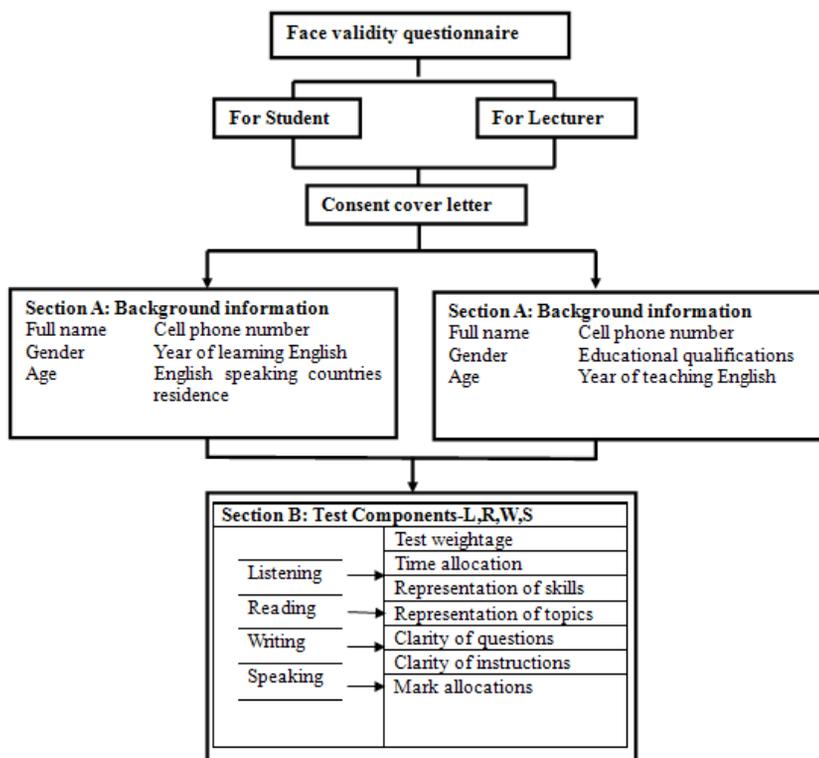


Figure 2. The adapted instrument framework

5.4. *The Institutional English Test*

The Institutional English language test consists of four English test components: Reading, Writing, Listening and Speaking.

Reading and Writing tests

Reading and Writing tests are taken together within 60 minutes. The Reading paper test consists of five parts with 55 questions while the Writing paper test has only two writing tasks.

1. Reading part 1: understanding messages
2. Reading part 2: three texts with questions
3. Reading part 3: long text with multiple choice questions
4. Reading part 4: text with multiple choice gaps
5. Reading part 5: text with gaps
6. Writing part 1: write a message
7. Writing part 2: write a story based on pictures

The Reading and Writing tests take 50% of the total score of the exams.

Listening Test

Students are required to complete 5 parts with 25 questions in the Listening paper test within 30 minutes. Each recording will be played twice

1. Listening part 1: pictures with multiple choice questions
2. Listening part 2: fill in a form
3. Listening part 3: multiple choice
4. Listening part 4: fill in a form
5. Listening part 5: longer conversation and matching

Each of the 25 listening questions scores 1 point. The Listening section is worth 25% of the total score of the exam.

Speaking Test

The IEST which is designed based on the common European Framework of Reference (level A2), has two parts which take 8-10 minutes. Generally, when students take the

speaking part of the IEST, they will do the examination with another candidate. The two of students will meet two examiners. One will do the talking while the other will take notes and assess their speaking.

Speaking part 1: A short Personal Information questions and answers exchange between candidate and the examiner.

Speaking part 2: The candidates will be given some cards with images/ideas or information on them and a card with some ideas for questions. After that one candidate will have to talk with the other candidate and ask or answer questions.

The speaking section is worth approximately 25% of the total score.

5.5. *Data collection and analysis procedures*

The set of data was collected through FVQ items given by 103 students and 20 English lecturers. This survey questionnaire was written in English, designed and adapted from several researchers (Cesur & Korsal , 2012; Dogru, 2013; Gonscar, 2008; Huong, 2001; Jaturapitakkul, 2013; Kucuk & Walters, 2009; Kuntasal, 2001; Moore, 2006 ; Pan, 1982 ; Wang, 2006) to get the opinions about the IET. During the survey, the instruction sheets were read out. After the participants finished filling out their background information questionnaire, they were asked to fill out the IET questionnaire. The results obtained from each question were administered, analyzed quantitatively and reported independently through the mean scores analyses in the SPSS program in order to investigate the perceptions of the validity of IET from both lecturers and students.

In order to establish the face validity of the IET, descriptive statistics analysis was made by computing the mean scores for each item in four components: Listening, Reading, Writing and Speaking in the students' and lecturers' questionnaire. Table 2 presents the interpretation of the mean scores:

Table 2. The interpretation of the mean scores

Mean	Option	Degree
4.5 - 5.0	Strong Agreement	Very high
3.5 - 4.4	Agreement	High
2.5 - 3.4	Neutral	Moderate
1.5 - 2.4	Disagreement	Low
1.0 - 1.4	Strong Disagreement	Very Low

Note: *VH =Very High, H=High, M=Moderate, L = Low, VL = Very Low

(Kucuk, 2007, p.65)

Table 2 shows the criteria of the mean scores adopted from Kucuk (2007). Five Likert-scale criteria were used to assess the degree in which the respondents agree to the face validity of the listening component. More precisely, the strongest agreement ranges from 4.5 to 5.0, followed closely by the agreement from 3.5 to 4.4 whereas undecided option covers 2.5 to 3.4. Last but not least, the disagreement starts from 1.5 to 2.4 and the strangles disagreement from 1.0 to 1.4.

In brief, the mean scores in Likert-scale criteria are used to measure the participant’s attitude by measuring the extent to which they agree or disagree with a particular question or statement

6. Findings and discussion

6.1. Participants

- 103 students participated in the survey, 63% of whom were females and 40% of them were males, aged between 18 and 22.

- 20 lecturers, 4 (20%) males and 16 (80%) females, who were teaching English. Their age ranges from 25 to 55.

6.2. Students’ perceptions

6.2.1. Students’ opinions on the IET

During the analysis procedure, descriptive statistics analysis was made by computing the mean scores for each item in for four components: Listening, Reading, Writing and Speaking in the students’ questionnaire in order to investigate the face validity of the IET. Table 3 shows mean scores for IET Components: Students’ perceptions:

Table 3. Mean scores for IET Components: Students’ perceptions (N=103)

Item	Listening Component			Reading Component			Writing Component			Speaking Component		
	Mean	SD	D	Mean	SD	D	Mean	SD	D	Mean	SD	D
Weightage	3.60	.664	H	3.74	.696	H	3.80	.667	H	3.76	.716	H
Time allocation	3.50	.765	H	3.69	.639	H	3.63	.656	H	3.77	.670	H
Skills	3.67	.687	H	3.64	.739	H	3.75	.706	H	3.76	.644	H
Topics	3.64	.904	H	3.61	.782	H	3.69	.764	H	3.71	.745	H
Questions	3.76	.846	H	3.73	.753	H	3.78	.824	H	3.83	.543	H
Instructions	3.84	.730	H	3.85	.567	H	3.81	.788	H	3.74	.750	H
Mark allocations	3.82	.788	H	3.71	.718	H	3.82	.686	H	3.78	.836	H

Note: *VH =Very High, H=High, M=Moderate, L = Low, VL = Very Low.

From Table 3, it can be noted that the mean scores for all items of four components are above 3.5, ranging from 3.5 to 3.8. In other words, four components of the IET generally seem to get much more satisfactory face validity from the students' opinions.

For Listening Component, the most positive comments are on the clarity of instructions with the highest mean score ($M = 3.84$, $SD = .730$). Meanwhile, the least positive comment is on time allocation ($M = 3.5$, $SD = .765$).

For Reading Component, the clarity of instructions ranks first with the highest mean score ($M = 3.85$, $SD = .567$) while the lowest mean score ($M = 3.61$, $SD = .782$) is accounted for the representation of topics.

For Writing Component, the most positive comments are on the mark allocations ($M = 3.82$, $SD = .686$) while the lowest mean score is for the time allocation ($M = 3.63$, $SD = .656$).

For Speaking Component, the most positive comments are on the clarity of

questions with the highest mean scores ($M = 3.83$, $SD = .543$) whereas the lowest mean score ($M = 3.71$, $SD = .745$) is on the representation of topics.

In short, the findings from face validity analysis in the students' questionnaire provide the evidence that each component in the IET gains satisfactory face validity from the students' opinions in terms of weightage, the time allocation, language skills, topics, questions, instructions and mark allocations. The IET, in other words, is perceived as a good test by students.

6.2.2. Student's mean scores of four IET components

The summary for face validity results on the IET: Listening, Reading, Writing and Speaking from the students' perceptions was made based on the analysis of the mean scores of four components. Table 4 shows the mean of 4 components: Listening, Reading, Writing and Speaking in the IET.

Table 4. Overall mean scores for IET's face validity results

Item (n=7)	Mean	SD	Degree	Rank
Listening	3.60	.764	H	4
Reading	3.71	.699	H	3
Writing	3.75	.728	H	2
Speaking	3.76	.701	H	1

Note: *VH =Very High, H=High, M=Moderate, L = Low, VL = Very Low

The mean scores show that the students' perceptions of four components in the IET are, on the whole, highly positive, ranging from 3.60 to 3.76. In particular, Speaking component ranks first with the highest mean score of 3.76 while Listening component ranks the lowest with 3.6. Therefore, with these highly positive comments, it can be claimed that the weightage and time allocation are appropriate to the students; therefore, they can distribute their answers and time

well according to the weight value of each component. Furthermore, the coverage of English skills and the representation of topics that students were taught during their English course were sufficiently represented in the IET components. Moreover, the students found no difficulties in understanding question items, instructions as well as identifying mark allocations for each part in each component; consequently, they can perform the IET better.

To sum up, the mean scores in the students' questionnaire show that each component in the IET gains satisfactory face validity from the students' perceptions in term of weightage, the time allocation, language skills, topics, questions, instructions and mark allocations. This means the IET is considered as a good in-house English language test by students.

6.3. Lecturers' perceptions

6.3.1. English language lecturers' opinions on the IET

Like students' questionnaire analysis for face validity of the IET, the descriptive statistics analysis was made by computing the mean scores for each item in for four components: Listening, Reading, Writing and Speaking in the lecturers' questionnaire. The interpretation of the mean scores is explained similarly to face validity results of the students' questionnaire analysis in Table 5 below:

Table 5. Mean scores for IET components Teachers' perceptions (N=20)

Item	Listening Component			Reading Component			Writing Component			Speaking Component		
	Mean	SD	D	Mean	SD	D	Mean	SD	D	Mean	SD	D
	Weightage	3.35	.745	M	3.55	.604	H	3.65	.670	H	3.95	.510
Time allocation	3.65	.670	H	3.85	.489	H	3.60	.680	H	3.90	.552	H
Skills	3.60	.680	H	3.64	.489	H	3.70	.571	H	4.10	.447	H
Topics	3.55	.686	H	3.65	.587	H	3.65	.587	H	4.10	.552	H
Questions	4.05	.686	H	4.05	.604	H	4.25	.444	H	4.15	.587	H
Instructions	4.10	.552	H	4.20	.410	H	4.30	.471	H	4.30	.470	H
Mark allocations	3.60	.753	H	3.55	.825	H	3.05	.887	M	3.10	.967	M

Note: *VH =Very High, H=High, M=Moderate, L = Low, VL = Very Low

Table 5 shows that for Listening Component, mean scores for all items are above 3.0, the mean degrees range from moderate to high face validity. The most positive comments are on the clarity of instructions with the highest mean scores (M= 4.10, SD = .552), meanwhile the least positive comment is for weightage (M=3.35, SD= .745).

For Reading Component, mean scores for all items are quite high, ranging from 3.5 to 4.2. Particularly, clarity of instructions ranks first with the highest mean score (M= 4.20, SD = 4.10) while the lowest mean score (M= 3.55, SD= .604) is accounted for the weightage. In other words, on the whole, the lecturers gave highly positive comments on the face validity of the reading component.

For Writing Component, mean scores for all items in the writing component are from

moderate to high, ranging from 3.05 to 4.30. In specific, the most positive comment is for clarity of instructions with the highest mean score (M= 4.30, SD= .471) whereas the lowest mean score (M=3.05, SD= .887) is for mark allocations.

For Speaking Component, mean scores for all items of the speaking component is similar to the writing component, ranging from moderate (M= 3.10) to high (M= 4.30). More precisely, the most positive comment is for clarity of questions with the highest mean scores (M= 4.30, SD= .470) whereas the lowest mean score (M=3.10, SD =.967) is for mark locations.

To sum up, lecturers' perceptions on four IET components are highly positive, ranging from 3.72 to 3.94. Among four components, Speaking ranks first with the

highest mean score of 3.94 while Listening Component ranks lowest with the mean of 3.72. This provides the evidence that each IET component gains satisfactory face validity from the lecturers' opinions in term of weightage, time allocation, skills, topics, questions, instructions and mark allocations. The IET, in other words, is also perceived as a good test from the lecturers' opinions.

6.3.2. Lecturers' mean scores of four IET components

The summary for face validity results on the IET: Listening, Reading, Writing and Speaking from the English lecturers' perceptions is made based on the analysis of the mean scores of four components in the IET. Table 6 shows the mean of four components: Listening, Reading, Writing and Speaking.

Table 6. Overall mean scores for Lecturers' IET face validity results

Item (n=7)	Mean	SD	Degree	Rank
Listening	3.72	.681	H	4
Reading	3.77	.572	H	2
Writing	3.74	.617	H	3
Speaking	3.94	.583	H	1

Note: *VH =Very High, H=High, M=Moderate, L = Low, VL = Very Low

Table 6 shows that the lecturers' perceptions on four IET components are highly positive, ranging from 3.72 to 3.94. Among four components, Speaking ranks first with the highest mean score of 3.94 while Listening Component ranks lowest with the mean of 3.72. This provides the evidence that each IET component gains satisfactory face validity from the lecturers' opinions in terms of weightage, time allocation, skills, topics, questions, instructions and mark allocations. The IET, in other words, is also perceived as a good test from the lecturers' opinions.

6.4. Comparison of students' and lecturers' perceptions about the IET

Due to the fact that all the questions in the second part of the students' questionnaire were parallel to the questions in the second part of the lecturers' questionnaire, and all of these questions were concerned with the face validity of the IET in term of weightage, time allocation, representation of skills, representation of topics, clarity of questions, clarity of instructions and mark allocations. Therefore, both students' and lecturers' perceptions of face validity of the IET are compared through the averaged mean scores of four components: Listening, Reading, Writing and Speaking in Table 7 below:

Table 7. Students' and lecturers' perceptions on the face validity of the IET

Component	Group	Mean	SD	Degree	Number
Listening	Students	3.60	.764	H	103
	Lecturers	3.72	.681	H	20
Reading	Students	3.71	.699	H	103
	Lecturers	3.77	.572	H	20
Writing	Students	3.75	.728	H	103
	Lecturers	3.74	.617	H	20
Speaking	Students	3.76	.701	H	103
	Lecturers	3.94	.583	H	20

Note: *VH =Very High, H=High, M=Moderate, L = Low, VL = Very Low

Table 7 shows that all mean scores between two groups possess high face validity. More specifically, the lecturers have more positive comments on Speaking and Reading components while students gain more satisfactory face validity on Speaking and Writing. Among these components, Speaking component achieves the most satisfactory face validity from both students and lecturers' opinions than the others. Fang and Faure (2010) showed that students tend to be more interested in developing communication skills than listening, reading and writing skills because they were increased the self-confidence of being of part in the classroom including answering the question, sharing the idea, and also presentation. Henning (1984) also emphasized that speaking lessons with different activities always help students and lecturers feel relaxed, more active in the learning and teaching process and at the same time make their learning and teaching more meaningful; thus, it could be explained that in this research students and lecturers feel much more comfortable during the speaking test than the others. Furthermore, speaking topics always relate to real-life situations which help students and lecturers feel freely when talking and sharing new ideas in the test.

In the light of descriptive statistical analysis above, it can be nutshell concluded that the IET is perceived as a good test from both students' and lecturers' perceptions.

7. Concluding remarks

Based on the evidence from the questionnaire surveys, it was found that the weightage and time allocation of the IET are appropriate to the students; therefore, they can distribute their answers and their time well according to the weight value of each component. Furthermore, the coverage of English skills and the representation of

topics that students were taught during their English course are sufficiently represented in the IET components. Moreover, the students find no difficulties in understanding question items, instructions as well as identifying mark allocations for each part in each component; consequently, they can perform the IET better. On the whole, the descriptive statistics on the face validity of the IET showed that both students and English lecturers possess satisfactory face validity of the IET. In other words, the IET, which is designed based on the Common European Framework of Reference, is perceived as a good English language test from the students and English lecturers.

The results are in line with similar earlier studies on face validity of in-house language tests such as Bachman (1990), Jaturapitakkul (2013), Kucuk (2007), Kuntasal (2001), Liauh (2011), and Nakamura (2006) that if the face validity of a language test is high, then it can be assumed that "the test itself will be successful, and the test-takers themselves may well perform as well as they might make test validity an important consideration in test use" (Bachman, 1990, p. 289). Therefore, it suggests that the perceptions of what the test measures from both students and educators, needs to be taken seriously as a potential factor which helps the students prepare themselves for the test; helps the lecturers improve the content of teaching and testing. Several researchers share the same view that face validity should be one of the intentions of test developers and should be integrated for test development with the crucial aim of promoting a better English education and learning at institutional level (Jaturapitakkul, 2013; Kucuk, 2007; Kuntasal, 2001; Liauh, 2011; Nakamura, 2006; Sato & Ikeda, 2015).

From the findings of this study, this could bring some implications and recommendations for both pedagogical and testing aspects. For example, lecturers should be concerned

with course objectives from the beginning. Additionally, they should teach and test in relation to course objectives or what is supposed to be measured. Likewise, by finding out that students need to know the question format before the test, lecturers should practice different question formats with students by revising previous tests or practicing some test-taking strategies. Also, providing quiet and comfortable environment helps students focus and improve their academic performance.

In a nutshell, it is hoped that the findings from the study have shed light on important factors which relate to the effect of the process of test preparation and test construction. Most importantly, it is further hoped that this study will be a contribution to the ongoing efforts to provide more validity evidence for in house English language tests.

References

- Alderson, C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press.
- Bachman, L. F. (1981). Languages for specific purposes: Program design and evaluation. *Formative evaluation in ESP program development*. Rowley, Massachusetts: Newbury House.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Palmer, A. (1996). The construct validation of self-ratings of communicative language ability. *Language Testing*, 6 (2), 13–20.
- Bachman, L. F., Davidson, F., & Milanovic, M. (1996). The use of test methods in the content analysis and design of EFL proficiency tests. *Language Testing*, 13(2), 125-150.
- Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, 27(3), 355–377.
- Brown, H. D. (2000). *Principles of language learning and teaching*. New York, US: Addison Wesley Longman.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, New York: Pearson Education.
- Campbell, D.T., & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Skokie, Illinois: Rand McNally.
- Campbell, K. (1996). *The world rushes to speak and write 'American' English*. London, UK: Oxford University Press.
- Carroll, J. B. (1968). *The psychology of language testing*. London, UK: Oxford University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Criper, C., & Davies, A. (1988). *The ELTS Validation Project Report*. London, UK: British Council and University of Cambridge Local Examinations Syndicate.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Philadelphia, US: Harcourt Brace Jovanovich College Publishers.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrical*, 16(3), 297-334.
- Crystal, D. (2000). *English as global language*. Cambridge, UK: Cambridge University Press.
- Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing*, 21 (2), 107–145.
- Cumming, A., & Alister. H., Berwick, R. (1988). *Validating in language testing*. Clevedon, England: Multilingual Matters.
- Davies, A. (1965). *Proficiency in English as a second language*. Birmingham, UK: University of Birmingham.
- Davies, A. (1978). Language testing: survey articles 1 and 2. *Language Teaching and Linguistic Abstracts*, 11, 145–59 and 215–31.
- Davies, A. (1984). Validating three tests of English language proficiency. *Language Testing*, 1(1), 50–69.
- Ebel, R. L., & Frisile, D. (1991). *Essentials of Educational Measurement (5th ed.)*. Englewood Cliffs, New Jersey: Prentice Hall.
- Elder, C. (1993). Language proficiency as a predictor of performance in teacher education. *Melbourne Papers in Language Testing*, 2, 68-85.
- Fang, T., & Faure, G. O. (2010). Chinese communication characteristics: A Yin Yang perspective. *International Journal of Intercultural Relations*, 35(3), 320–333. DOI: 10.1016/j.ijintrel.2010.06.005.
- Ferguson, G., & White, E. (1998). A small scale study in predictive validity. *Melbourne Papers in Language Testing*, 7, 15 -63.
- Fernandes, H. J. X. (1984). *Testing and Measurement*. Jakarta, Indonesia: National Education Planning.
- Fives, H., & Didonato-Barnes, N. (2013). Using a table

- of specification to improve teacher-constructed traditional tests: an experimental design. *Assessment in Education: Principles, Policy and Practice*, 21(1), 90-108.
- Fulcher, G. (1997). An English language placement test: issues in reliability and validity. *Language Testing*, 14(2), 113-139.
- Glass, G. V., & Hopkins, K. D. (1984). *Test Bank for Statistical Methods in Education & Psychology, Second Edition*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Graham, J. G. (1987). English language proficiency and prediction of academic success. *TESOL Quarterly*, 21(3), 505-521.
- Grissom, R. J., & Kim, J.J. (2012). *Effect sizes for research: Univariate and multivariate applications (2nd ed.)*. New York, U.S: Routledge.
- Harrison, A. (1983). *Communicative testing: Jam tomorrow?* London, UK: Academic Press.
- Heaton, J. (1988). *Writing English language tests*. London, UK: Longman.
- Henning, G. (1984). Current Developments in Language Testing. *Language Testing in Asia*, 1, 237-241.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Cambridge, UK: Newbury House.
- Herbert, C. (2004). Integrating Testing with Teaching. *Practical Assessment, Research & Evaluation*, 1(6), 3-8.
- Hill, K., Storch, N., & Lynch, B. (1999). A comparison of IELTS and TOEFL as predictors of academic success. *International English Language Testing System (IELTS) Research Report*, 2, 62-73.
- Hughes, A. (1989). *Testing for Language Teachers Second Edition*. UK: Cambridge University Press.
- Khalifa, H., & Weir, C. (2009). Examining reading: Research and practice in assessing second language reading. *Studies in Language Testing*, 29. Cambridge, UK: Cambridge University Press.
- Kindeya, N.T. (2002). *The content validity of the Ethiopian general secondary education certificate English examination* (Unpublished master's thesis). Addis Ababa University, Ethiopia.
- Kinyua, K., & Okunya, L.O. (2014). Validity and reliability of teacher-made tests: Case study of year 11 physics in Nyahururu District of Kenya. *African Educational Research Journal*, 2(2), 61-7.
- Kluitmann, S. (2008). *Testing English as a foreign language: Two EFL-Tests Used in Germany* (Unpublished master's thesis). University of Albert-Ludwig, Germany.
- Kucuk, P. (2007). *The relationship among face validity, reliability and predictive validity of University English Foreign Language preparatory school achievement tests* (Unpublished master thesis). Bilkent University, Turkey.
- Kuntasal, I. (2001). *Perceptions of teachers and testers of achievement tests prepared by testers in the Department of Basic English at Middle East Technical University* (Unpublished master's thesis). Bilkent University, Turkey.
- Kuroki, K. (1996). *Achievement testing: A final achievement test model for Japanese junior high school students* (Unpublished master thesis). Bilkent University, Turkey.
- Le, V.C., & Barnard, R.(2009). Curricular innovation behind closed classroom doors. *Prospect*, 24(2), 20-33.
- Lee, Y. J., & Greene, J. (2007). The predictive validity of an ESL placement test: A mixed methods approach. *Journal of Mixed Methods Research*, 1(4), 366-389.
- Liao, Y. (2002). Issues of validity and reliability in Second Language Performance Assessment. *Language Testing*, 4(2), 2-5.
- Samad, A.A., Rahman, S. Z. S., & Yahya, S.N. (2008). Refining English language tests for university admission: A Malaysian example. *Asian Journal University of Education*, 4(1), 57-68.
- To, T.T.H. (2000). *Investigating the predictive validity of IELTS among Vietnamese students* (Unpublished doctoral thesis). La Trobe University, Australia.
- Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The writing strategies of graduate research students in the social sciences, *Higher Education*, 27, 379-392.
- Weir, C. (1993). *Understanding and developing language tests*. New York, US: Prentice Hall.
- Weir, C. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke, England: Palgrave Macmillan.
- Weir, C. J., Chan, S. H. C., & Nakatsuhara, F. (2013). Examining the criterion-related validity of the GEPT advanced reading and writing tests: comparing GEPT with IELTS and real-life academic performance. *LTTC-GEPT Research Report*, 1, 1-43.

GIÁ TRỊ BỀ NGOÀI CỦA BÀI KIỂM TRA TIẾNG ANH NỘI BỘ THEO KHUNG CHƯƠNG TRÌNH CHÂU ÂU TẠI MỘT TRƯỜNG ĐẠI HỌC CÔNG LẬP Ở VIỆT NAM

Nông Thị Hiền Hương

Trường Đại học Nông lâm Thái Nguyên

Tân Thịnh, Thái Nguyên, Việt Nam

Tóm tắt: Trong kiểm tra đánh giá ngôn ngữ, giá trị bề ngoài của một bài kiểm tra được sử dụng bởi những người học và có thể được coi là loại hiệu lực kiểm tra được thảo luận phổ biến nhất vì nó chủ yếu được xử lý với câu hỏi liệu một bài kiểm tra có đo được những gì người học dạy trong khoá học. Do đó, nghiên cứu này điều tra nhận thức của các giảng viên tiếng Anh và sinh viên về các bài kiểm tra tiếng Anh nội bộ theo khung chương trình Châu Âu được tổ chức tại một trường đại học công lập ở Việt Nam. Một cuộc khảo sát với 103 sinh viên và 20 giảng viên tiếng Anh từ chương trình giáo dục đại học công lập đã được thực hiện. Một bảng câu hỏi với 7 mối quan tâm chính - trọng số, phân bổ thời gian, kỹ năng ngôn ngữ, chủ đề, mục câu hỏi, hướng dẫn và phân bổ nhân hiệu đã được sử dụng để thu thập dữ liệu. Tất cả các câu trả lời được phân tích thông qua thống kê mô tả. Kết quả cho thấy theo ý kiến của cả sinh viên và giảng viên, bài kiểm tra tiếng Anh nội bộ theo khung chương trình Châu Âu có giá trị bề ngoài hợp lệ, do đó, bài kiểm tra tiếng Anh nội bộ theo khung chương trình Châu Âu được coi là một bài kiểm tra tốt để đo lường khả năng tiếng Anh của sinh viên.

Từ khoá: kiểm tra ngôn ngữ, độ giá trị của bài kiểm tra, kiểm tra độ giá trị, giá trị bề ngoài, xác nhận kiểm tra

APPENDICES

Appendix 1: IET Face Validity Questionnaire (Student)

Date:

Dear Participant,

I would like to invite you to kindly participate in this study by completing the enclosed questionnaire. The survey should take about 10-15 minutes to complete. Your participation in this study is voluntary and you are free to withdraw your participation from this study at any time. There are no risks associated with participating in this study. All responses in the survey will be recorded anonymously and used for the purpose of this research only.

If you agree to participate in this study, please respond to the items in the questionnaire. Your honest opinions are very crucial for this research.

If you have any questions regarding the survey or the research in general, please feel free to contact Nong Thi Hien Huong at hhuong04052002@yahoo.com

Thank you for your cooperation and time.

Sincerely,

Nong Thi Hien Huong

Email: hhuong04052002@yahoo.com

Cellphone number: +84.984.888.345

This questionnaire aims to gather your opinion about the IET conducted at a public university in Vietnam. Please kindly respond to all the items.

Section A : Background Information							
For Section A, please fill up the space provided with relevant information.							
1.	Full Name	:					
2.	Matric No	:					
3.	Gender	:	<input type="checkbox"/> Male	<input type="checkbox"/> Female			
4.	Age	:					
5.	Email	:					
6.	Cellphone number	:					
7.	How many years have you been learning English	:					
8.	Have you lived in English speaking country/countries for more than 3 months during a single stay?	:	<input type="checkbox"/> Yes	<input type="checkbox"/> No			
9.	If yes to Item 8		What country/countries?	:			
			How long in total?	:			
Section B: IET Components							
This section aims to gather information on the IET components comprises of Listening Test, Reading Test, Writing Test and Speaking Test.							
Please circle only ONE of the options given below.							
Strongly disagree - SD			Neutral - N				
Disagree - D			Agree - A				
			Strongly agree - SA				
Question 10- 16: Listening Test							
10	Weightage for the listening test was appropriate		SD	D	N	A	SA
11	Time allocation for the listening test was sufficient		SD	D	N	A	SA
12	Listening skills taught were sufficiently represented in the listening test		SD	D	N	A	SA
13	Topics taught were sufficiently represented in the listening test		SD	D	N	A	SA
14	Questions in the listening test were clear		SD	D	N	A	SA
15	Instructions explaining what to do in each section of the listening test were clear		SD	D	N	A	SA
16	Marks allocated for each section of the listening test were stated clearly		SD	D	N	A	SA

Question 17- 23: Reading Test						
17	Weightage for the reading test was appropriate	SD	D	N	A	SA
18	Time allocation for the reading test was sufficient	SD	D	N	A	SA
19	Reading skills taught were sufficiently represented in the reading test	SD	D	N	A	SA
20	Topics taught were sufficiently represented in the reading test	SD	D	N	A	SA
21	Questions in the reading test were clear	SD	D	N	A	SA
22	Instructions explaining what to do in each section of the reading test were clear	SD	D	N	A	SA
23	Marks allocated for each section of the reading test were stated-clearly	SD	D	N	A	SA
Question 24- 30: Writing Test						
24	Weightage for the writing test was appropriate	SD	D	N	A	SA
25	Time allocation for the writing test was sufficient	SD	D	N	A	SA
26	Writing skills taught were sufficiently represented in the writing test	SD	D	N	A	SA
27	Topics taught were sufficiently represented in the writing test	SD	D	N	A	SA
28	Questions in the writing test were clear	SD	D	N	A	SA
29	Instructions explaining what to do in each section of the writing test were clear	SD	D	N	A	SA
30	Marks allocated for each section of the writing test were stated-clearly	SD	D	N	A	SA
Question 31- 37: Speaking Test						
31	Weightage for the speaking test was appropriate	SD	D	N	A	SA
32	Time allocation for the speaking test was sufficient	SD	D	N	A	SA
33	Speaking skills taught were sufficiently represented in the speaking test	SD	D	N	A	SA
34	Topics taught were sufficiently represented in the speaking test	SD	D	N	A	SA
35	Questions in the speaking test were clear	SD	D	N	A	SA
36	Instructions explaining what to do in each section of the speaking test were clear	SD	D	N	A	SA
37	Marks allocated for each section of the speaking test were stated-clearly	SD	D	N	A	SA

Thank you for your co-operation

Appendix 2: IET Face Validity Questionnaire (English Lecturer)

This questionnaire aims to gather your opinion about the IET conducted at a public university in Vietnam. Please kindly respond to all the items.

Section A : Background Information			
For Section A, please fill up the space provided with relevant information.			
1.	Full Name	:	

2.	Gender	:		Male		Female
3.	Age	:				
4.	Email	:				
5.	Cellphone number	:				
6.	What is your highest academic qualification?		• B.A			
			• M.A			
			• Ph.D.			
7.	How long have you been teaching English at your university?		• Less than 5 years			
			• 5 to 10 years			
			• 11 to 15 years			
			• More than 15 years			

Section B: IET Components

This section aims to gather information on the IET components which comprise of Listening Test, Reading Test, Writing Test and Speaking Test.

From questions 1 to 28, please circle only **ONE** of the options given.

Strongly disagree - SD		Neutral - N
Disagree - D		Agree - A
		Strongly agree - SA

Question 1- 7: Listening Test

1	Weightage for the listening test was appropriate	SD	D	N	A	SA
2	Time allocation for the listening test was sufficient	SD	D	N	A	SA
3	Listening skills taught were sufficiently represented in the listening test	SD	D	N	A	SA
4	Topics taught were sufficiently represented in the listening test	SD	D	N	A	SA
5	Questions in the listening test were clear	SD	D	N	A	SA
6	Instructions explaining what to do in each section of the listening test were clear	SD	D	N	A	SA
7	Marks allocated for each section of the listening test were stated clearly	SD	D	N	A	SA

Question 8- 14: Reading Test

8	Weightage for the reading test was appropriate	SD	D	N	A	SA
9	Time allocation for the reading test was sufficient	SD	D	N	A	SA
10	Reading skills taught were sufficiently represented in the reading test	SD	D	N	A	SA
11	Topics taught were sufficiently represented in the reading test	SD	D	N	A	SA
12	Questions in the reading test were clear	SD	D	N	A	SA

13	Instructions explaining what to do in each section of the reading test were clear	SD	D	N	A	SA
14	Marks allocated for each section of the reading test were stated clearly	SD	D	N	A	SA
Question 15- 21: Writing Test						
15	Weightage for the writing test was appropriate	SD	D	N	A	SA
16	Time allocation for the writing test was sufficient	SD	D	N	A	SA
17	Writing skills taught were sufficiently represented in the writing test	SD	D	N	A	SA
18	Topics taught were sufficiently represented in the writing test	SD	D	N	A	SA
19	Questions in the writing test were clear	SD	D	N	A	SA
20	Instructions explaining what to do in each section of the writing test were clear	SD	D	N	A	SA
21	Marks allocated for each section of the writing test were stated clearly	SD	D	N	A	SA
Question 22- 28: Speaking Test						
22	Weightage for the speaking test was appropriate	SD	D	N	A	SA
23	Time allocation for the speaking test was sufficient	SD	D	N	A	SA
24	Speaking skills taught were sufficiently represented in the speaking test	SD	D	N	A	SA
25	Topics taught were sufficiently represented in the speaking test	SD	D	N	A	SA
26	Questions in the speaking test were clear	SD	D	N	A	SA
27	Instructions explaining what to do in each section of the speaking test were clear	SD	D	N	A	SA
28	Marks allocated for each section of the speaking test were stated clearly	SD	D	N	A	SA

Thank you for your co-operation