

Student and teacher perceptions of an English language test for university admission: Understanding its face validity

Received: *July 25, 2022*

Revised: *September 6, 2022*

Accepted: *December 18, 2022*

Sasithorn Jackson

Faculty of Liberal Arts, Mahidol University, Thailand

sasithorn.jac@student.mahidol.edu

Abstract

The perception of test users should be an important consideration in creating a test because there are some issues that only test users can respond to. Furthermore, the perception that teachers have of a test is also an important consideration and it may affect their instructional practices. Therefore, the research questions were (1) to what extent do student perceptions agree with teacher perceptions of the abilities that are involved in answering test items? And (2) what types of test items create differences between student and teacher perceptions? Sixty undergraduate students, 19 high-school teachers, and 19 university teachers participated in an online questionnaire. Twenty-two of all participants responded to the follow-up interview. Findings from the questionnaire showed that students and teachers agreed that a test item did aim to assess the ability mentioned in the questionnaire but they were uncertain if the test could assess their/ students' ability to write, speak, and communicate in English. Moreover, the interview results showed that the English GAT test had washback on students' test preparation and high-school teachers' teaching or/and assessment but not on university teachers. It is hoped that the results of this study will raise the awareness of those who deal with the language test, especially the communicative test, and guide them to consider the potential influence of the test's face validity upon testing outcomes as they relate to real-world abilities.

Keywords: face validity, test-taker perception, English test, university entrance examination, washback

1. Introduction

For many reasons, including administrative cost and time, and subjective scoring, among others, a number of tests assessing English communication administered in Thailand, including the General Aptitude Test (GAT) still use a multiple-choice format which is considered an indirect test. Due to the specificity of the testing format, it is often the case that drills and rote memorization are considered more conducive to test success than practicing the communicative language, itself (Prapphal, 2008). Oral communication normally involves interaction among people; therefore, one could easily argue that testing for oral communication skills using a multiple-choice format test is a method that obviously lacks face validity¹. Face validity ought not to be a separate quality of tests because it can reveal flaws in the test development that may have been overlooked by test developers. Moreover, face validity can also impact the targeted assessment results (Xie, 2011). According to Watanabe (2004), there is a widespread belief that gaining entrance to and studying at prestigious universities is the winning ticket to a great career and an easy life after graduation. Thus, results of university entrance examination are seriously important and unlikely to be ignored or disregarded. In order to achieve the intended washback² on the content of teaching and learning, test developers need to consider face validity as judged by the test users; in other words, test developers must consider their test users' point of view (Sato & Ikeda, 2015; Watanabe, 2004) because test users are influenced by the skills or abilities they believe the test is measuring (Sato & Ikeda, 2015).

As of this writing, however, there is no information regarding the perception of test users towards GAT available. Therefore, a compelling reason to follow this line of research exists in order to attain a better understanding of students' and teachers' perception of GAT. The research questions are:

- 1) To what extent do student perceptions agree with English teacher perceptions of the abilities that are involved in answering test items?
- 2) What types of test items create differences between student and English teacher perceptions?

¹ Face validity: the degree that test users judge that a test is appropriate to the assessment construct and objectives (Nevo, 1985)

² Washback: influence from testing on teaching and learning (Alderson & Wall, 1993; Hughes, 1989)

2. Literature review

2.1 Relationship between face validity and washback

Face validity is a judgment of the test content by non-testing experts such as students and teachers (Alderson, Clapman & Wall, 1995). The other term for face validity could be test-takers' attitudes towards a language test (Fan, 2014). Mosier (1947) said a practical language test should not only have statistical validity, but it also should appear practical, relevant, and applicable to the test users.

Washback is an effect that tests can have upon teaching and learning (Hughes, 2003). Washback can be a positive or a negative effect (Davies, Brown, & Elder, 1999). Hughes (2003) suggested how to improve positive washback effects: 1) Use direct testing. When the abilities that students need to develop are directly tested, then test preparation must consist of practices aiming to develop those abilities. 2) Ensure that the test is understood by students and teachers. They need to understand what the test requires of them so that they can benefit from the positive washback the test should be designed to promote (Hughes, 2003). 3) The test's specifications, and sample items must be available to everyone dealing with the test preparation (Hughes, 2003; Yan, Thirakunkovit, Kauper, & Ginther, 2016).

Watanabe (2004) emphasized that not only the test, itself, can cause washback but also the face validity of the test can cause washback, especially from a student's point of view. Yan et al. (2016) stated that test-takers' perceptions can support test developers to better evaluate and identify improvements for their test.

Traditional paper-and-pencil tests seem to lack face validity and therefore do not intentionally produce positive washback. Struyven, Dochy, and Janssens (2005) found a strong relationship between face validity and washback upon learning through their meta-analysis of 36 empirical studies. They found that many students perceived traditional paper-and-pencil tests as random and irrelevant. Consequently, students purposefully learned only what they needed to pass such tests, with no intention of maintaining their knowledge or abilities through, for example, life-long learning activities. The researchers emphasized that what students perceive about the test (its face validity) seems to have a substantial washback on students' learning.

More support comes from Sato and Ikeda (2015). According to their analysis, Korean test-takers did not perceive that the abilities measured in the College Scholastic Ability Test (CSAT) were what the test committees intended to measure. Instead of the test evaluating writing abilities in English, the test takers perceived the test's goal as aiming to assess their reading abilities. The primary reason underlying such a finding is likely to be the format of the

test. The test tasks only involved placing given sentences and paragraphs in the correct order, summarizing a paragraph into one sentence, and placing a given sentence in the correct place of the passage, all of which are accomplished in a multiple-choice format. Consequently, instead of practicing writing in preparation for the test, students could practice mainly reading skills or focus on multiple-choice test-taking strategies.

On the other hand, the General English Proficiency Test (GEPT) administered in Taiwan accomplished the purpose of upgrading Taiwanese English proficiency and motivating English learning there (Vongpumivitch, 2012). The researcher found that both groups of participants (61.7% of the students and 71% of the non-students) reported that the GEPT was fair and reliable. Moreover, the participants reported that they needed a test like GEPT to show how much English they knew (mean of the students = 4.85 and mean of the non-students = 4.73). This finding supported what Chan, Schmitt, DeShon, Clause, & Delbridge (1997) and Hughes (2003) stated. The results of a test with high face validity can better reflect test-takers' real abilities. The other reason for this test to have accomplished its goals might be found in the formatting of the test; it was a direct test. Test takers had to demonstrate their abilities through listening, speaking, and writing.

Zhou and Yoshitomi (2019) also found that the test takers who perceived the test as having higher validity tended to give more importance to the test, and they then increased their effort. In short, a test with high face validity can increase test-takers' effort to perform well on the test. Watanabe (2004) believes that face validity as understood from the test-users' point of view is an important test characteristic to support positive washback. In addition, The International Language Testing Association (2021) and Shohamy (1993) stated that test developers have a particular duty to improve the quality of their language tests. This suggests that test developers ought to consider face validity as an important test characteristic when creating an assessment and intentionally create positive washback.

2.2 English admission tests in Thailand

Due to the large scale of Thai university entrance examinations, multiple-choice format tests are mainly used for the purpose of reliable and practical scoring. The washback effects of such test contents and formats are observably apparent (Imsa-ard, 2020; Lunrasri, 2014). For example, in many schools, senior high-school students were prepared with the materials that had the same content and format as those used for the university entrance examinations (Imsa-ard, 2020; Lunrasri, 2014; Prapphal, 2008), with special emphasis given to this type of preparation in the final semester of high school (Prapphal, 2008).

Most teachers agreed that the English ONET (Ordinary National Educational Test) preparation had a washback effect upon their teaching. Teachers taught to what was being tested, used ONET-like materials, created school tests based on the previous years' ONET, devoted more time for test preparation, and assigned homework relating to the ONET. In addition, both students and teachers believed that the contents in the test were different from the contents in the textbooks (Lunrasri, 2014). That is likely to be the major reason why the teachers taught the way they did.

Moreover, the teachers thought that they were driven by the English ONET test rather than by the textbook or the curriculum. They did not believe that their students' ONET scores reflected their students' abilities to use language in real-life situations. The teachers also agreed that the students' scores can be raised by teaching to the test. Moreover, 79% of the teachers thought that the test needed to be comprised of other features such as listening, speaking, and writing. The researcher concluded that an indirect test cannot assess students' communicative abilities (Imsa-ard, 2020).

Research studies have shown that how students and teachers perceive a test in terms of its contents and formats (face validity) can, indeed, affect washback of the test. In Thailand, researchers have investigated students' and teachers' perceptions towards the English ONET and found that the test did have a great amount of washback upon them (Imsa-ard, 2020; Lunrasri, 2014). Still, test-users' perceptions towards the English GAT (General Aptitude Test) are unknown. Therefore, this study aims to explore students' and teachers' perceptions of such a test.

3. Methods

3.1 Context of the study

In Thailand, senior high school students must take a test called GAT (General Aptitude Test) for national university admission. According to test information on the website, NEST (2020) the English GAT aims to measure ability to communicate in English. It consists of four sections: speaking, vocabulary, reading comprehension, and structure and writing. The test consists of 60 items in total, and students are given 90 minutes to complete.

3.2 Participants

The participants in this study consisted of three groups: 60 first-year undergraduate students, 19 high-school English teachers, and 19 university English teachers. The students were from one public university in Nakhonpathom. This university was chosen because students from all over Thailand have come to study here so the sample could possibly represent

Thai students who had experienced taking the GAT test. The high-school teachers were from Thai schools which were some of the oldest schools in Bangkok and metropolitan areas of Thailand (i.e., Samutprakan, Nakhonpathom, and Nonthaburi). The criteria for choosing the schools were 1) the schools were extra-large sized (having 2,500 students or more) because the students at extra-large sized schools passed the ONET English with the highest scores on average among all schools (NIETS, 2020). 2) Next, among schools that passed with the highest ONET English average scores, only those that were under the Secondary Educational Service Area Office or the Office of the Private Education Commission would be considered (NIETS, 2020). 3) Schools to be selected would have been in service for more than 50 years as an indicator of educational quality. The university teachers were from Thai public universities in which some of them are among the oldest universities in Thailand. The criterion for choosing the universities was that the universities would have been in service for more than 70 years as an indicator of educational quality.

All participants completed the online questionnaire. Then, ten of the students, five of the high-school teachers, and seven of the university teachers participated in the follow-up interview. All participants were selected on a voluntary basis.

3.3 Instruments

The instruments in this study consisted of a questionnaire, and a semi-structured interview.

3.3.1 Questionnaire

The online questionnaire consisted of three parts:

Part 1: demographic information

Part 2: 17 test items and agreement rating (Likert-scale)

An original GAT test of 60 items was sampled to produce a sample test to be administered for this study. Like the original test, the sample test still consisted of four parts: speaking, vocabulary, reading, and structure and writing parts. Items were selected by considering all of the same abilities involved in completing the original test and covering all of the same formats as in the original test. The sample GAT included 5 items for the speaking part, 3 items for the vocabulary part, 7 items for the reading part, and 2 items for the structure and writing part. They were written in English as in the original. Both students and teachers received the same sample GAT test. Each item asked the participants to rate their level of agreement with different statements that they were presented with. Possible responses were: strongly agree, agree, not sure, disagree, and strongly disagree. Data was also collected from written

comments for “If you were not sure or disagreed with the stated ability, please specify the ability you think the item is aiming to assess.”

Part 3: three questions eliciting the participants’ overall perceptions on the test

The instructions in the questionnaire were written all in Thai to avoid language difficulty. No time limit was set for the completion of the questionnaire thus, allowing participants to deliberate upon the abilities being measured.

The questionnaire was assessed in terms of content validity by three experts. The experts were native Thai speakers who have proper training and are qualified in making such assessments (Ph.D.). In addition, the questionnaire was assessed in terms of internal reliability using Cronbach’s Alpha. Every part of the questionnaire received Cronbach’s Alpha score greater than 0.70 which was considered as acceptable (Sijtsma, 2009).

3.3.2 Semi-structured Interview

The semi-structured interview contained four open-ended questions to elicit students’ and teachers’ responses about their perceptions towards the English GAT test. Both students and teachers were asked question items 1-2. One additional question was used for teacher and student group. Furthermore, all interviewees who gave the test items a rating of “Not Sure”, “Disagree”, or “Strongly Disagree” were asked “You were not sure/disagreed/strongly disagreed with the stated ability being tested in item number ... Can you explain why you chose that response?” The reason for using the semi-structured interview is that it offers a balance between flexibility in the order of the questions and focus on the topic. The interview was conducted in Thai to avoid language difficulty and misunderstanding.

3.4 Procedure

The first set of data was collected through an online questionnaire (Google Form) in the first semester of the academic year 2021. For the second set of data, 22 participants had an online interview with the researcher. All participants were assured that they had right not to participate in this study. Even when they were participating, they had right to decline answering any questions that they felt uncomfortable with and the right to withdraw from the research at any time. They were also assured that their withdrawal would have no effects on their studies or career.

The data was analyzed with the Statistical Package for the Social Sciences (SPSS) version 18.0. The percentage, mean, and S.D. were calculated to investigate what extent student perceptions and English teacher perceptions of the abilities involved in answering test items agree (17 items) as well as their perceptions towards the test overall (3 items). Frequency count was used for written comments. In addition, the items that seemed to induce the gap between

student and teacher perception were further investigated. For interview data, content analysis was used. The interview data was first transcribed. Then the researcher read the transcriptions several times to summarize and analyze the key findings based on responses of the interviewees to each question.

4. Results

4.1 Perceptions of the abilities that are involved

The data in this section was collected from 17 Likert-scale items (98 participants). The graphs below indicate the participants' level of agreement: SA = strongly agree; A = agree; NS = not sure; D = disagree; and SD = strongly disagree. The graphs also distinguish between S = student participants and T = teacher participants.

Table 1 showed that students and English teachers agreed that an item aimed to measure a mentioned ability (M = 4.1, S.D. = 0.4 for students and M = 4.1, S.D. = 0.6 for teachers). However, the students and teachers were uncertain that the test could measure writing, speaking, and communication abilities (M = 3.1, S.D. = 0.9 for students and M = 3.4, S.D. = 0.8 for teachers).

Table 1 Summary of students and English teacher perceptions

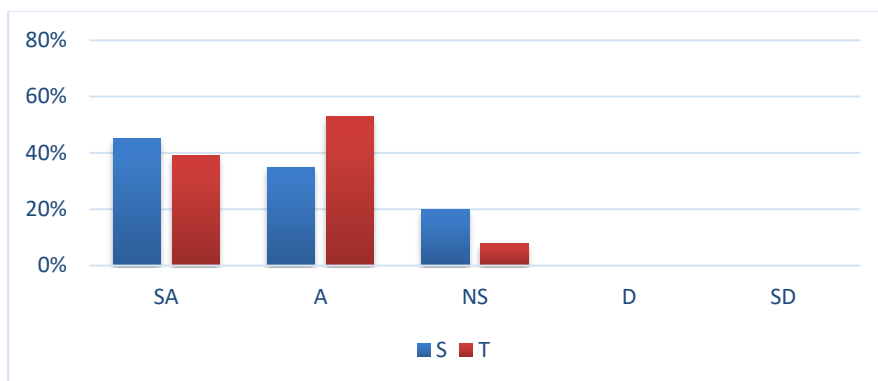
Category	Students			Teachers		
	M	S.D.	Interpretation	M	S.D.	Interpretation
Agree that an item measures a mentioned ability (17 items)	4.1	0.4	Agree	4.1	0.6	Agree
The extent to the test overall measures a mentioned ability (3 items)	3.1	0.9	Not sure	3.4	0.8	Not sure

The number of students who agreed that each item (total of 17 items) aimed to assess a mentioned ability is 80% on average. The test items that received less than 80% of student agreement are items number 4, 9, 11, 13, and 16. The number of teachers who agreed that each item (total of 17 items) aimed to assess a mentioned ability is 85% on average. The test items that received less than 85% of teacher agreement are items number 4, 5, 12, 13, 16, and 17.

1) A: I've been very busy. So much work to finish before the exams.

B: _____. You'll be surprised how quickly it'll all be over.

- a. So far so good
- b. Good for you
- c. Hang in there
- d. Get back on track



Most students and almost every teacher agreed that item 1 aimed to assess “Knowledge of vocabulary/ expressions and appropriateness of meaning”.

Items 2-3:

Susan: What are you doing?

Supa: I'm studying for the final exams.

Susan: 2 ! I haven't done anything yet, I'm not in the mood.

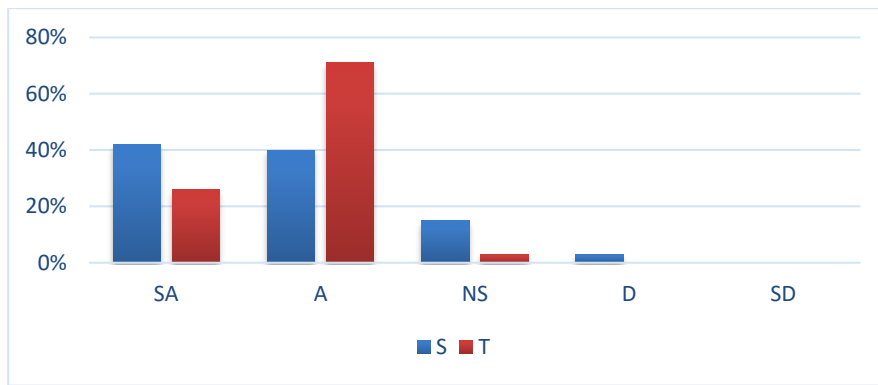
Supa: Well, the exams aren't that far off. 3 I won't have enough time to prepare.

Susan: Do you think the exams will be difficult?

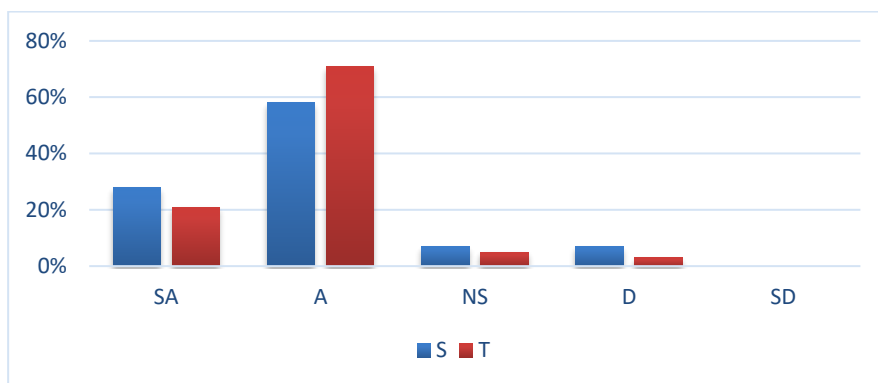
Supa: Of course. That's why I'm preparing for them.

Susan: Well then. I think it's a good idea to study for the exams now. Thanks for reminding me.

- 2) a. How clever I am
- b. How happy you are
 - c. What a good student
 - d. What a nice occasion



- 3) a. I'm sorry
 b. I'm afraid
 c. I can't imagine
 d. I can't believe



Most students and almost every teacher agreed that items 2 and 3 aimed to assess “Understand coherence of the dialogue” and “Understand cohesion of the questioned sentence”, respectively.

Items 4-5

Paula: Linda, did you manage to get hold of Jane after work yesterday?

Linda: No, I'm sorry, Paula. 4 last night. And this morning I tried again, but nobody answered.

Paula: Don't you have her cell phone number?

Linda: I'm sorry, I don't. Have you got it?

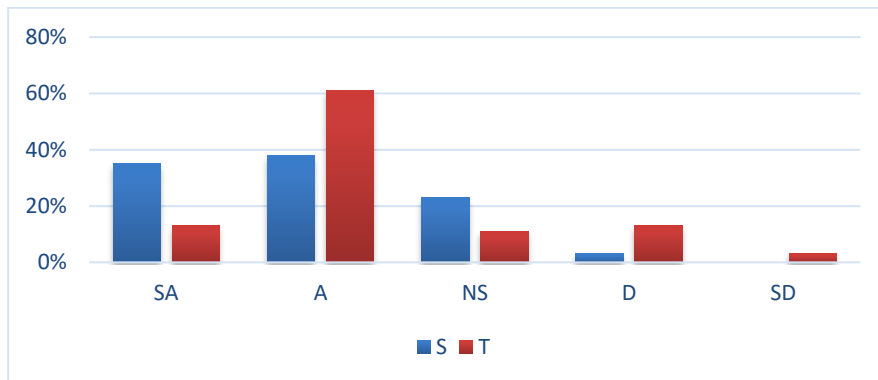
Paula: Of course I do. You should have called me last night.

Linda: Well, I thought you were busy preparing for the meeting, and I didn't want to disturb you. 5?

Paula: The meeting starts in 10 minutes. That's out of the question. She'll never get here in time.

- 4) a. I didn't put her through

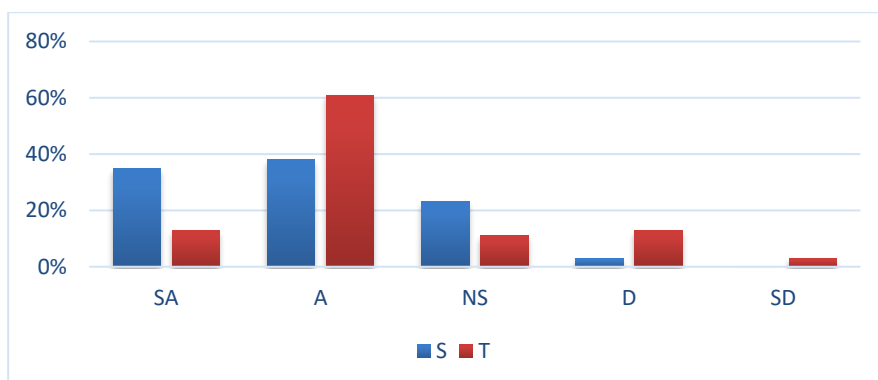
- b. I didn't have her number
- c. Her line was always busy
- d. Her phone wouldn't answer



A below average number of students and teachers agreed that item 4 aimed to assess “Knowledge of everyday expressions and appropriateness of meaning”. Among the 22 interviewees, two students and two teachers gave it a rating of “Not Sure”. Two (a student and a teacher) of them believed that it was not about everyday expressions. Moreover, the other teacher believed that students need to know more than everyday expressions as he said

“I was uncertain because students need to know vocabulary and collocation such as “get hold of”. I think that students also need to use their imagination and cognitive ability to infer the meaning. That is why I answered “Not Sure”.”

- 5) a. Should I call her now
- b. When is the meeting
 - c. How soon is the meeting
 - d. Would you take her call



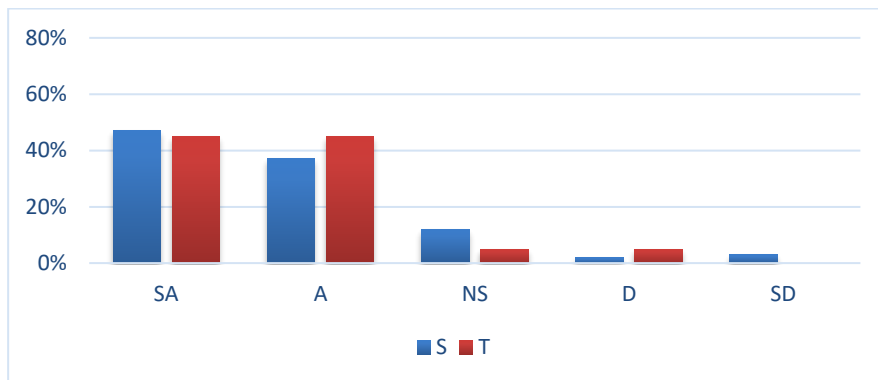
The students and the teachers have quite different opinions towards item 5. The majority of students agreed that the item aimed to assess “Ability to choose a question whose meaning is appropriate to the context” but a below average number of teachers agreed so. Among 22

interviewees, three teachers gave it a rating of “Not Sure” and one teacher gave it a rating of “Strongly Disagree”. Two teachers believed that the context was unclear. This is one of their opinions:

“I think the correct answers can be options a, b, or c depending on an individual’s style of talk. Students can get confused if the test is not clear. And if students choose option b, it means wrong answer, right? Then students will be judged that they do not have ability to choose a question whose meaning is appropriate to the context.”

Item 6: Three of the words in each group relate to one another in their meaning. Choose the word that does NOT belong.

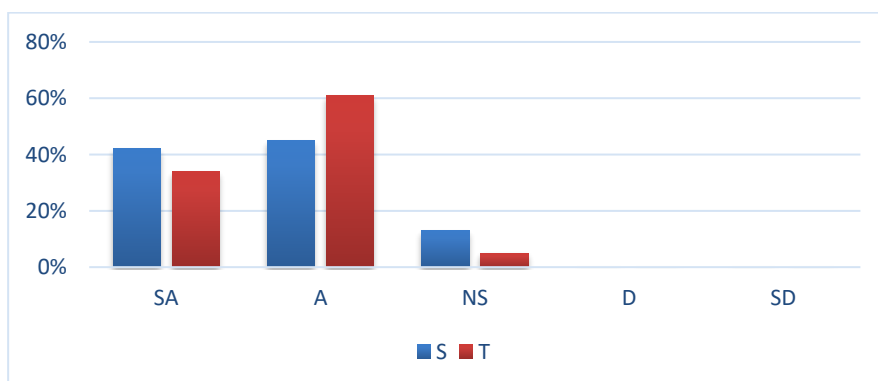
a. grab b. seize c. snatch d. rise



Item 7: Choose the alternative which has the same meaning as the underlined word in the given sentence.

Apply some moisturizer on your skin when it gets dry.

- Students should apply what they have learned in their real life.
- I had to apply the brake very quickly.
- I will apply a plaster on the cut to your finger.
- My sister wants to apply for a new job.

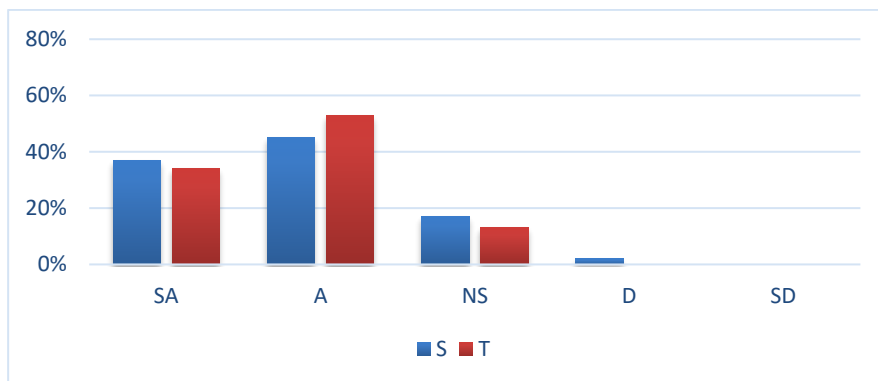


Most students and almost every teacher agreed that items 6 and 7 aimed to assess “Knowledge of vocabulary (without context)” and “Knowledge of vocabulary and appropriateness of meaning”, respectively.

Item 8: Choose the best alternative to make the sentence meaningful.

Valuable _____ of oil and gas lies hidden in rocks on the _____ where they are tapped by drilling down through those rocks.

- a. reservoirs – seabed
- b. collections - sea wall
- c. resources - sea level
- d. assets – seaboard



Most students and teachers agreed that item 8 aimed to assess “Knowledge of vocabulary and appropriateness of meaning”. Among 22 interviewees, two teachers gave it a rating of “Not Sure”. One of them thought that the context was not precise enough that she could choose the right vocabulary. She said

“The context could not help me to select between “seabed” and “sea level” and between “reservoirs” and “resources”. If the context is good, it should help us decide what to select. But I have to decide based on my knowledge, instead. The context did not help me.”

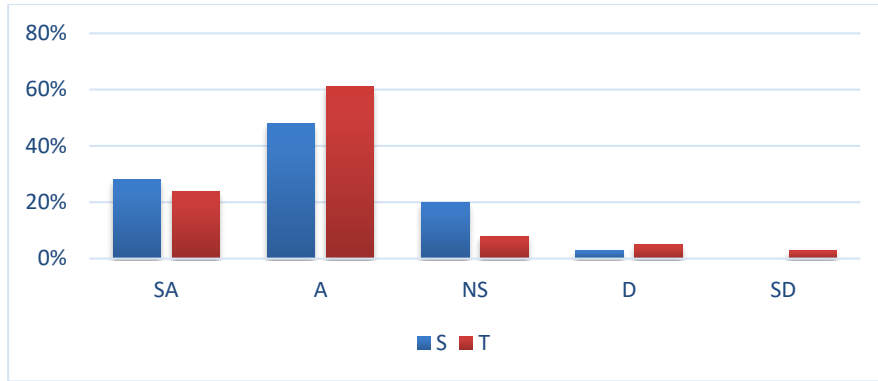
The other believed that some of the vocabulary items in the options are not related to students’ real life such as sea wall and seabed.

Item 9:

What can be inferred from the comic strip?



- a. The girl doesn't say no.
- b. The boy doesn't say no.
- c. The boy doesn't like the girl.
- d. The girl doesn't like the boy.



The students and the teachers have slightly different opinions towards item 9. A below average number of students agreed that the item aimed to assess “Understand coherence of the dialogue and appropriateness of meaning” but the majority of teachers agreed so. Among 22 interviewees, two students and one teacher gave it a rating of “Not Sure”. One of the students said she did not know the meaning of “Gag me with a spoon” and if it could be used in daily-life conversation. The teacher believed that students who do not have enough experience with the language might have difficulty answering this question. She said:

“I think in order to answer this item; students need to use their own experience at a specific level. If readers do not have background knowledge and only look at the context, they might not be able to answer this item. I feel that this item takes world knowledge more than literal texts to answer because the literal text does not help much about inferred meaning. If students do not have experience, they might answer option a. Another thing that helps is test-taking strategies to get rid of impossible options. The question asked about “infer” so option “a” cannot be the right answer. In conclusion, only context is not enough.”

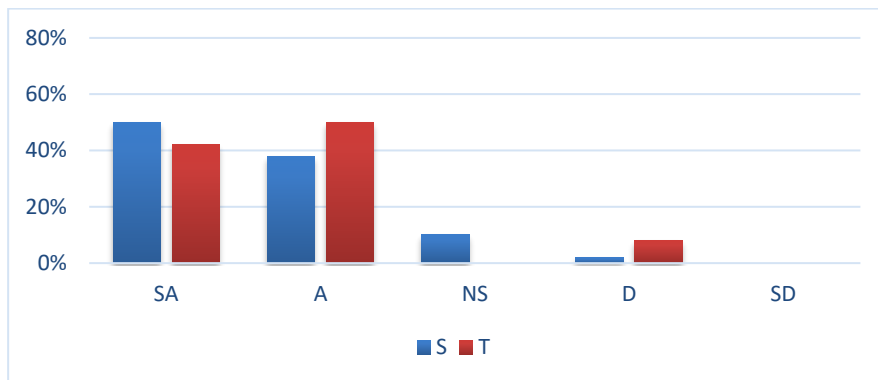
Item 10:

Smoke-Free

- a. You are free to smoke in this area.
- b. There is no air pollution in this area.

c. You are not allowed to smoke in this area.

d. Free cigarettes are available in this area.



Most students and almost every teacher agreed that item 10 aimed to assess “Ability to read signs and understand the meaning”.

Passage 1: Item 11-13

(1) My family and I have watched with interest the swallows that, over the years, have sought out a nest attached to the wall underneath the small terrace of our house. They come and go, and we look forward to their return each spring.

(2) One day I noticed that the nest had fallen over with the small occupants - six fat fledglings covered in good plumage - still inside. My eight- year-old son arranged an emergency shelter: he put the babies inside a cardboard box, along with bits of cotton wool and a few pieces of the clay from the destroyed nest. When my husband returned home that evening, we considered how we could build a new nest for the swallows.

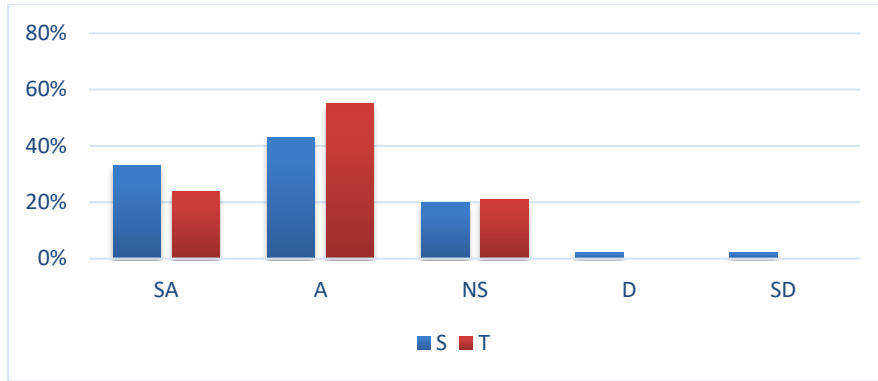
(3) After some animated discussion and without hope of success, we decided to put the baby birds in another deeper cardboard box, together with the clay and soft cotton wool. With a few nails in the wall and a few holes in the improvised nest, the box was attached to the wall with its little occupants inside. We worried that our handling of the tiny chicks had left the odor of humans on them. Would the parents, when they returned, become suspicious, or even reject their young ones?

(4) So we waited anxiously. With dawn came the sound of the parents busily going about the job of feeding six hungry birds! They looked after their brood until the fledglings became self-sufficient; only then did they dedicate themselves to fixing up the new nest we had built.

(5) Some 22 years later, this nest is still sought out by successive generations of swallows who come and depart- the source of renewed happiness and joy for us.

11) What is the first paragraph mainly about?

- a. A family activity
- b. A shelter for swallows
- c. Environment protection
- d. The small terrace of a house



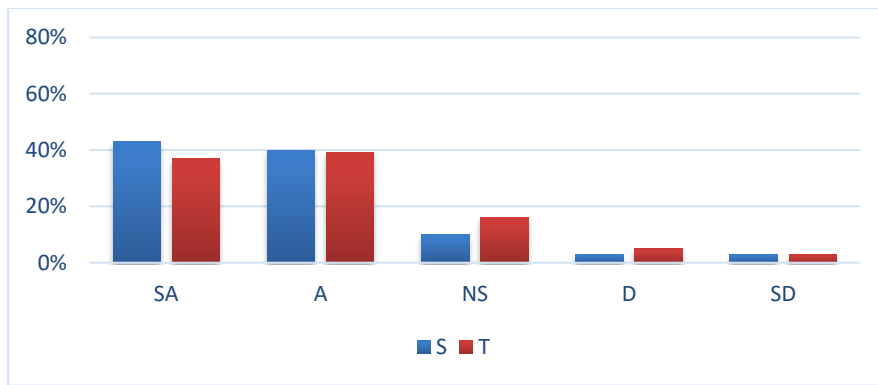
A below average number of students agreed that item 11 aimed to assess “Understand the main idea of the paragraph” but the majority of teachers agreed so. Among the 22 interviewees, two students and two teachers gave it a rating of “Not Sure”. Two (a student and a teacher) of them thought that there was more than one correct answer. The teacher gave her opinion as

“Maybe the answer is “Family Activity”. I am not sure if the correct answer should be option a or b so I had to read the whole story. It is a little ambiguous. After I read the first paragraph, I thought it should be “Family Activity”. But after I read the whole story, I thought it was rather about “A Shelter for Swallows”. I think the first paragraph should be the main idea but the main idea of this first paragraph is not that clear.”

Another teacher interviewee said that he could choose the correct answer by using test-taking strategies. He also said that after he read the paragraph, the main idea that came to him was not among any of the given options.

12) What was NOT used to make a new nest for the baby swallows?

- a. A terrace
- b. A cardboard box
- c. Cotton wool
- d. Clay



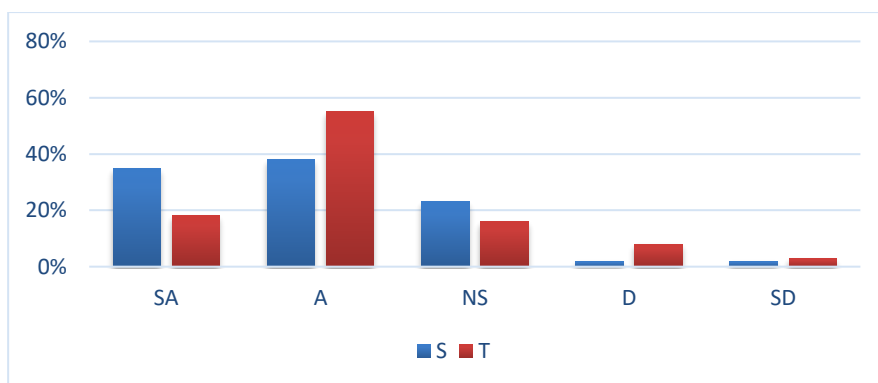
The students and the teachers have slightly different opinions towards item 12. Most students agreed that the item aimed to assess “Understand details of the paragraph” but a below average number of teachers agreed so. Correspondingly, two teachers who rated “Disagree” wrote their comments as

I can guess the answer without reading the story.

The options are obviously matched with the story.

13) What can be concluded from the passage?

- a. Environmental protection is necessary.
- b. A friendship across species is possible.
- c. Swallows are man's best friends.
- d. Life is not easy for swallows.



A below average number of students and teachers agreed that item 13 aimed to assess “Understand the story conclusion”. Correspondingly, among 22 interviewees, two students and a teacher gave it a rating of “Not Sure” and one teacher gave it a “Strongly Disagree”. The two students believed that there was more than one correct answer. In addition, the two teachers stated that the conclusion statement could have been better said. This is one of the teachers’ opinions

“If I could conclude it myself, what I conclude would not be any of the given options. I feel that the story and the conclusion do not quite go together. If I

can set the conclusion, it will not be these (in the options). I think the given options and the story should be more coherent than these.”

Passage 2: Item 14-15

(1) A recent study found that resident doctors know when they are exhausted, upset or overwhelmed and when they are, they are far more likely to admit they made an error. What they are admitting is that sleepiness is not the only factor. Above and beyond that is general distress and mental fatigue, researchers reported in the *Journal of the American Medical Association*.

(2) “While fatigue is important, there is this whole domain of distress beyond fatigue that also demands attention,” Colin West of the Mayo Clinic in Rochester, Minnesota, who led the study, said in an interview. The findings may help point to ways to better reduce the burdens on resident doctors, known in some countries as junior doctors, and in turn prevent mistakes. “I think this is going to have an impact on healthcare reform,” Dr. West said. “We need to put resources into training and medicine to control work hours and maintain physician well-being.”

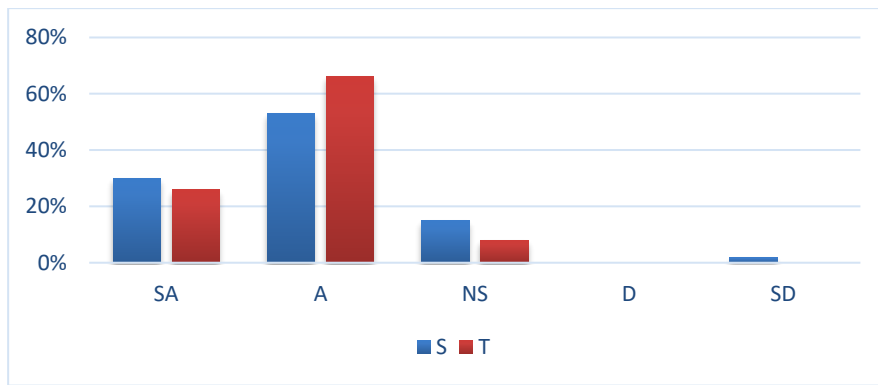
(3) The US Institute of Medicine reported in 1999 that between 48,000 and 98,000 Americans die each year from preventable medical errors ranging from drug overdoses to infections caught in the hospital. And doctors, unions and other experts have been **clamoring** to cut the work hours of residents, who are required to work at least 50 hours a week at most training hospitals.

(4) Dr. West and colleagues surveyed 356 residents at 163 medical schools globally. They found that 39% reported making at least one major medical error during the study period, and these doctors were also more likely to say they were sleepy, fatigued, or stressed “What we have shown in this data is that fatigue is important, but it's only part of the issue and previous studies have not paid much attention to the distress factors,” Dr West said “Over the course of their training and the course of their career everyone has made a major error. Everyone tries to be perfect but no one is. He said only limiting work hours may not be enough.

(5) Teaching hospitals across the United States have moved to limit residents’ work weeks to 40 hours to reduce fatigue-related errors. It is estimated that **this** could cost US\$171 million to \$487 million a year, depending on whether additional residents are hired or substitute providers are brought in.

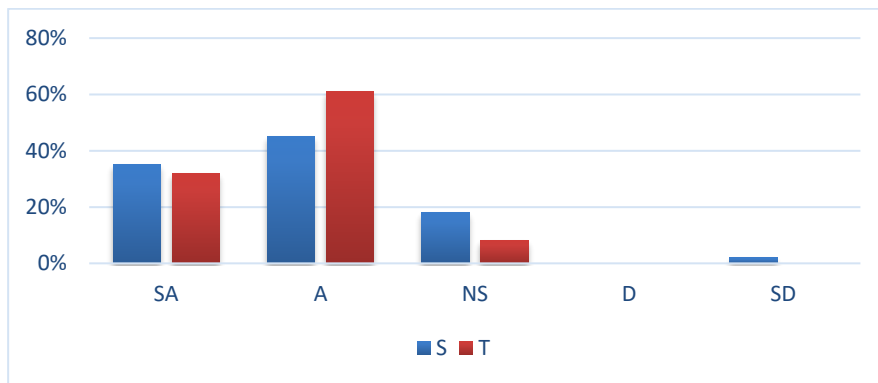
14) Which word is closest in meaning to "clamoring" (paragraph 3)?

- | | |
|---------------|----------------|
| a. Demanding | b. Rejecting |
| c. Announcing | d. Encouraging |



15) What does "this" (paragraph 5) refer to?

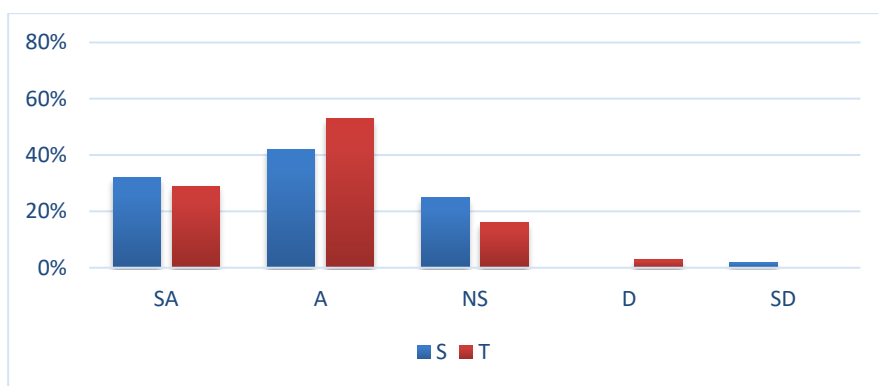
- a. Trying to be perfect
- b. Being sleepy, fatigued, or stressed
- c. Limiting residents' work hours
- d. Hiring additional residents



Most students and almost every teacher agreed that items 14 and 15 aimed to assess “Knowledge of vocabulary and appropriateness of meaning” and “Understand the story detail and the relation of sentences”, respectively. Correspondingly, none of the interviewees gave these two items a rating of “Not Sure”, “Disagree”, or “Strongly Disagree”.

Item 16: **Read the following statements and choose the underlined part that is grammatically wrong.**

If buildings (a) sprang up suddenly out of the ground (b) like mushrooms, their rooftops would be covered (c) into a layer of (d) soil and plants.



A below average number of students and teachers agreed that item 16 aimed to assess “Knowledge of structure and English grammar”. Among 22 interviewees, a student and a teacher gave it a rating of “Not Sure”. These are the teacher and the student’s opinions, respectively.

“I thought it was assessing rather collocation or preposition, because I saw “covered into”. Then I knew that it was incorrect.”

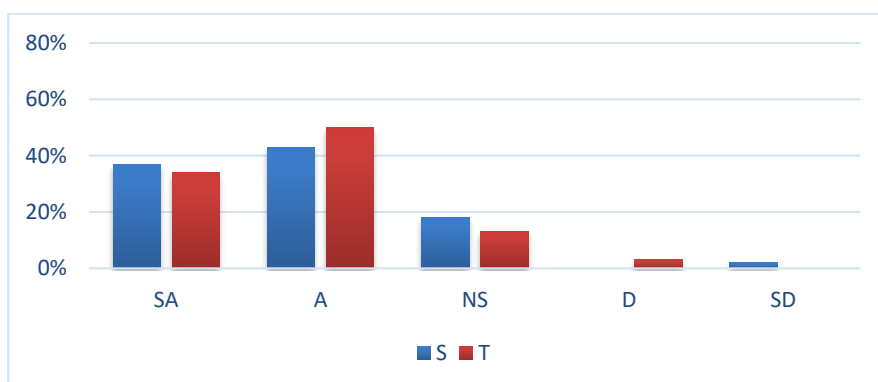
“It may also assess vocabulary. Not only assess grammar or structure. It may be other things.”

Item 17:

Choose the best answer to complete the sentence.

The government spent much more in the last quarter than planned _____ it spent considerably less in the previous one.

- a. whereas
- b. before
- c. until
- d. when



Most students and teachers agreed that item 17 aimed to assess “Understand the meaning and the relation of the sentences and able to choose the correct conjunction”. Not surprisingly, none of the interviewees gave it a rating of “Not Sure”, “Disagree”, or “Strongly Disagree”. Only one teacher wrote his comment stating that the meaning of this sentence is unclear.

4.2 Perceptions towards the test overall

The data in this section was collected from 3 Likert-scale items about overall opinions (98 participants). The data was also collected from the follow-up interview. All interviewees were asked “Could you explain why you gave such rating for that this test could measure writing/speaking/communication abilities?”

About half of students and teachers believed that this test can assess their/students’ ability to write in English (See table 2).

Table 2 Students and teacher perceptions towards the test overall

Statement	The Most (%)		Much (%)		Not sure (%)		Little (%)		The Least (%)	
	S	T	S	T	S	T	S	T	S	T
To what extent do you think that this test can assess your/students’ ability to write in English	5 (8)	5 (13)	26 (43)	19 (50)	15 (25)	4 (11)	12 (20)	6 (16)	2 (3)	4 (11)
To what extent do you think that this test can assess your/students’ ability to speak in English	1 (2)	0	21 (35)	18 (47)	11 (18)	10 (26)	21 (35)	7 (18)	6 (10)	3 (8)
To what extent do you think that this test can assess your/students’ ability to communicate in English	4 (7)	5 (13)	26 (43)	23 (61)	15 (25)	4 (11)	10 (17)	5 (13)	5 (8)	1 (3)

Among 22 interviewees, four students said they wanted to write instead of taking a multiple-choice format for writing test. Three teachers supported the students’ opinions that a multiple-choice test might not be able to accurately measure writing abilities. This is one of the teachers’ opinions

“I cannot see how this test can help improve students’ writing. “Error” can barely assess writing. If wanting to assess writing, ones should require students to really write.”

On the other hand, two teachers believed that this test can assess writing. One of them specified “Error Identification” as being able to do so. In conclusion, the majority of the interviewees believed that a multiple-choice test could not truly assess writing abilities. In addition, five of them suggested a writing, direct test instead of a multiple-choice test.

Less than half of students and teachers believed that this test can assess their/students’ abilities to speak in English (See table 2). Among 22 interviewees, two students stated that they wanted to speak to assess their speaking abilities. In addition, four teachers believed that there is no fixed answers for real communication. Here are two of the teachers’ opinions

“Communication does not always have to follow the given options (There are more than one way to respond to ones’ interlocutors). I also think about spoken grammar.”

“I am not sure because: 1) each student has an unequal background, so we do not know how much students have in terms of English abilities. How much vocabulary they have, how much they are exposed to English. The test is paper-based so we cannot see their interactions based on their different background. 2) In some dialogues, there are no fixed answers. We can communicate in many ways with the same result. Therefore, this test can assess speaking abilities at a “medium” level. It is not $1+1 = 2$. It is communication and communication is dynamic. And we do not know each student’s background therefore, assessing speaking abilities with a paper-based test is limited with fixed answers. It is not very effective. There is no fixed formula in communication. It has suprasegmental, body language, feedback, and other strategies. So in order to say that this student is good or weak, this test can do it at a “medium” level.”

However, a student and two teachers believed that the conversation part of the test could, in fact, assess speaking abilities. In conclusion, the majority of the interviewees seemed to be dissatisfied with the multiple-choice, paper-based test assessing speaking abilities. The main reason they gave is that in real communication there are no fixed answers.

Half of the students and 74% of the teachers believed that this test could assess students’ abilities to communicate in English (See table 2). Among 22 interviewees, one student believed that if one can communicate in English, then they should be able to do well on this test. Another

student and one teacher believed that this test could assess communication because of the conversation and reading parts. In addition, one teacher praised this test in that it covers a variety of texts including texts that can be understood literally and also ones that readers need to use their abilities to infer the meaning of the test item. On the other hand, three teachers did not consider reading and/or grammar as part of communication. They seemed to understand that communication is about speaking. Here is one of their opinions:

“I see this test as focusing on reading, vocabulary, and grammar more than on communication. If one wants to assess communication, the questions should be open-ended and then ask students to write their answers. If it is a fixed, multiple-choice test, I think that it can barely assess communication then.”

In addition, a student supported the above opinion that the format of the test needed to be changed. The student stated that communicative tests ought to be more than paper-based exams. However, another teacher understood that speaking and listening had to be adapted into a paper-based exam. One teacher emphasized testing a variety of ways and levels in communication as well as in reading passages because they can also affect the communicative assessment. He said:

“There are many ways, many levels in communication. So I responded “Much” but not “The Most”. Does the GAT measure communication in daily life or academia? Spoken or written language? There are also many kinds of reading passages.”

In conclusion, about half of the interviewees seemed to believe that this test can assess abilities to communicate in English. However, two interviewees believed that to assess communication, the test needs to be more than a paper-based, multiple-choice format exam. Surprisingly, one misperception about communication, thinking it did not include reading and grammar, was noticed in the interviews.

On average, both students and English teachers were uncertain if the test could measure writing, speaking, and communication abilities (See table 1). This result is in line with the written comments: “Please identify the communicative abilities you think this test aims to assess.” Writing, speaking, and communication abilities were reported to be perceived as measured less than the abilities of reading, vocabulary, and grammar (See table 3).

Table 3 Summary of the written comments

Abilities participants mentioned	Frequency	%
Reading	45	30
Vocabulary and expressions	36	24
Grammar and structure	26	17
Writing	18	12
Conversation and speaking	17	11
Communication	10	6
Total	152	100

4.3 The items that seem to create differences in student and teacher perceptions

The issues that are found to create differences in student and teacher perceptions are categorized as in Table 4. The test items to be discussed in this section are 4, 5, 6, 8, 9, and 16.

Table 4 Issues creating difference in student and teacher perceptions

Issues	Test items
Insufficient context	4, 5, 8
Inauthentic language use	6
Outdated language	9
Options were not absolutely correct / clear	5, 9
Requirement of advanced level of vocabulary / world knowledge	4, 5, 8, 9, 16

It is obvious that the context of items 4 and 5 is not relevant to students' daily life. Therefore, students had to use their advanced level of world knowledge in order to understand the situation. Moreover, students needed to acquire an advanced level of vocabulary to know the meaning of "get hold of". Last, the options in item 5 were not absolutely correct. Options a, b, and c can be the question for "The meeting starts in 10 minutes." but options b and c are incorrect because of "That's out of the question. She'll never get here in time."

Item 6 is concerned about inauthentic language use. The vocabularies used here are not purposeful and contextual. For item 8, it is obvious that three vocabularies which are seabed, sea wall, and seaboard are not relevant to the student-life context. Students needed to have an advanced level of vocabulary knowledge in order to choose the right word. In addition, the given context was not enough for test takers to decide between the last two options.

Item 9 is concerned with outdated slang language, “Gag me with a spoon.” One issue with this item is the requirement of an advanced level of world knowledge. That expression was used briefly in the 1970-1980s as representative of California English, and specifically the San Fernando Valley in Los Angeles. Due to its obscure context, students needed to use their world knowledge to interpret and understand the situation happening in the cartoons. Lastly, option d was not absolutely correct over option a.

Besides grammar and structure knowledge, item 16 was considered as including the advanced level of vocabulary knowledge especially collocations such as “cover with” and “spring up”.

4.4 Washback upon students and teachers

Among ten student interviewees, more than half of them reported that the GAT test affected their opportunity to pursue their intended field in universities. They said if they got low scores on the GAT, they might not get a university seat in the fields they wanted to pursue. For this reason, five of them reported that they had to practice using the previous years’ tests from books or the internet, or even take a special course at a tutoring school.

Overall, the GAT had washback on students’ test preparation including practicing using the previous years’ tests or extra studying at a tutoring school. None of the students prepared for their test by using English naturally.

Among the 12 teacher interviewees, all high-school teachers admitted that the GAT affects their teaching. The major reasons they gave were that senior high-school students need to continue their studies in universities and the GAT is one of the national admission exams. Three teachers revealed that: 1) at their schools there was a new English subject whose purpose is to prepare students for the national exams including the GAT and/or 2) they had to find the books or test examples that related to the GAT then compare their contents in order to prepare their students for the test. However, two of the teachers expressed concern about differences in contents between schoolbooks and the test. The two teachers also agreed that teachers in schools had to teach communication but also had to prepare students for the national tests. The instructions of teaching communication and preparing for the national tests did not go in the same direction. On the other hand, the GAT did not have washback upon university teachers. One of them said that the test that might affect students in universities is the TOEIC. She also emphasized that, personally, she preferred her students to be able to demonstrate their speaking or communication abilities rather than taking a paper-based test.

Overall, the GAT had washback on high-school teachers but not on university teachers. The high-school teachers admitted that the test affected their teaching and/or assessment. For example, they used test samples as teaching materials.

5. Discussion and Conclusions

The results of this research revealed that the students and English teachers who participated agreed, on average, that a test item did aim to measure a mentioned ability. Several issues are discussed here in order to create awareness of the matter of face validity which can affect learning and teaching of those issues.

The test items that received agreement from the majority of students and teachers are those that are in the contexts related to students' life experience, for example, test preparation and conversing about their lives. On the other hand, the items that a below-average number of students and/or teachers agreed with are in the contexts beyond those of students' usual life experience, for example, items from the context of a workplace situation. This testing practice, in fact, contradicts what Mosier (1947) suggested. A practical language test ought to be relevant to the test users.

Vocabulary assessed in the test ought to be frequently-used words such as "apply". According to several word lists, including Oxford Learner's Word Lists (2022) and Cambridge English (2022), "apply" is considered a high frequency word that English learners need to know. In addition, archaic language or dated slang expressions, such as "Gag me with a spoon", need to be avoided. According to Corpus of Contemporary American English (2022), "Gag me with a spoon" had a frequency of 14 from the years 1193 to 2017 which is considered relatively insignificant. Hughes (2003) suggested using published reference material to check frequently-used words when selecting vocabulary in large-scale exams. Using common words and language that people regularly use supports students' learning and will be useful for them in real communication, not just in test preparation. Besides the issue of frequency, vocabulary to be tested ought to be in context in order to create meaningful communication because people do learn vocabulary when it is integrated with social, cultural, and life contexts (Chen & Li, 2010). Vocabulary out of context can be ambiguous because words often have multiple meanings and context is the key to unlocking the meaning. In other words, a vocabulary test item without context can be considered as inauthentic and therefore cannot truly measure students' communicative abilities.

In addition, test response options also need to be carefully created. The correct answers need to be the unambiguously correct choices over the distractors as Brown (2005) suggested.

Otherwise the item might not be able to accurately measure students' real abilities and they may develop a negative attitude towards an ambiguous test item or towards the test overall with the result of removing the desired positive washback effect that a well-made test can create.

To support fairness for every student, an item that requires world knowledge to reveal its intended meaning needs to be carefully selected. In other words, the required world knowledge ought to be commonly acquired by the students so that a test item can treat every student as fairly as possible (Hughes, 2003).

In conclusion, test developers need to carefully create a test item in order to avoid items consisting of irrelevant-to-student contexts, outdated language, and infrequent vocabulary items. In addition, the context ought to be substantially sufficient for students to choose the right answer which needs to be unambiguously correct over other distractors.

For the test overall, it is found that the students and teachers were uncertain if the test could measure writing, speaking, and communication abilities. These results are in line with the written comments. Reading, vocabulary, and grammar were mentioned as abilities this test most aimed to assess and less so were writing, speaking, and communication abilities. The reason underlying this perception might be the format of the test which is multiple-choice. The students and the teachers thought that they did not have to write or speak; they only chose the answers. This is in line with Sato and Ikeda's (2015) study. They found that the students perceived the test item differently from the test committees. Instead of assessing writing, students considered it as a reading test. The main reason accounting for that is the test format, which is multiple-choice. In addition, a number of participants in this study believed that a direct test can better assess abilities to write and/or speak in English than a multiple-choice test can. Ideally, a direct test can accurately measure such abilities. In reality, using a direct test means having to deal with more complex administration and scoring issues. Therefore, understanding the test objectives and specifications, as well as the necessity of using a multiple-choice format for the national test, should encourage students and teachers to directly learn and teach English abilities instead of solely practicing "learn for the test" and "teach to the test" activities. Teach real abilities, not rote activities.

Another reason could be a misunderstanding about communicative competence which is excluding reading and/or grammar from communication. The participants who reported to believe so tended to have a negative attitude towards the test. Therefore, understanding the definition and scope of communicative competence can help students and teachers to see the clear path of what they need to learn and to teach. This is supported by Hughes (2003) and Watanabe (2004). Watanabe (2004) stated that student and teacher perceptions towards the test

(face validity) affect their learning and teaching. Hughes (2003) claimed that in order to promote the positive washback that the test should intend to create, students and teachers need to understand what such a test requires of them. If reading and grammar had been included in communication, communication might have been mentioned more than it was. Therefore, the suggestion above might be able to help students and teachers know the scope of what they need to learn or to teach as well as prevent them from solely engaging in “learn for the test” and “teach to the test” activities. Moreover, this deep understanding about the definition and scope of communication could prevent students and teachers from having negative attitudes towards such a test.

It is obvious that the GAT had washback upon students and high-school teachers. The students and teachers seemed to only engage in “learn for the test” and “teach to the test” activities, respectively. The perceptions they held towards the test and the test format are likely to be the reason why they did so. The teachers in this study revealed similar practices to the teachers in the studies of Imsa-ard (2020) and Lunrasri (2014). In order to support positive washback and also to generate reliable and practical test scoring in this national-scale exam, besides the same GAT format, test developers may add a part where students must listen to spoken English in order to choose the correct response. The added format would likely lead students and teachers to improve their learning and teaching at least by practicing listening. Semi-direct testing would promote moving away from practicing to the test and towards practicing for the exam by using natural English in communication as a normal part of classroom instruction.

6. Limitations and Recommendation

Due to the drawback of the Likert scale, “Not Sure” cannot be precisely interpreted since the participant was uncertain about the given statement. For example, some participants who did not have strong feelings for a given statement may choose “Not Sure” despite their actual agreement. For this study, such an issue was partly mitigated by asking the interviewees to explain why they gave a rating of “Not Sure” for a specific item.

The version of the English GAT that is available on the NIETS official website is considered relatively out-of-date. There is a substantial difference in the formats between the version used in this study and the latest version. Therefore, there could be some aspects in the latest test that were not addressed in this study.

The number and the gender of participants in the student and teacher groups were unbalanced due to the difficulty created by the Covid-19 pandemic. When schools and

universities resume normal activities, future researchers can potentially replicate this present study with a larger sample size and a group balanced by number and gender to increase relevance to the real world.

Regarding the participants of this study, the researcher conducted the study with only students and English teachers. There could be further research conducted with other stakeholders such as curriculum developers, school principals, and parents to understand their perceptions of the test (face validity) in order to support the positive washback the test has the latent potential to promote.

From this study, it is found that students and teachers were uncertain if this test could assess writing, speaking, and communicative abilities. In addition, they believed that this test aimed to mainly assess reading skills, vocabulary, and grammar knowledge. Future researchers can examine if such perceptions (face validity) have a significant relationship with what and how students learn and what and how teachers teach.

References

- Alderson, J. C., & Wall, D. (1993). Does washback exist?. *Applied linguistics*, 14(2), 115-129.
- Alderson, J. C., Clapman, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge University Press.
- Brown, J. D. (2005). *Testing in language programs: a comprehensive guide to English language assessment*. McGraw-Hill College.
- Cambridge English. (2022, May 2). Cambridge English Vocabulary List. <https://www.cambridgeenglish.org/images/84669-pet-vocabulary-list.pdf>
- Chan, D., Schmitt N., DeShon, R. P., Clause C. S., & Delbridge K. (1997). Reactions to Cognitive Ability Tests: The Relationships Between Race, Test Performance, Face Validity Perceptions, and Test-Taking Motivation. *Journal of Applied Psychology*, 82(2), 300-310.
- Chen, C. M., & Li, Y. L. (2010). Personalised context-aware ubiquitous learning system for supporting effective English vocabulary learning. *Interactive Learning Environments*, 18(4), 341-364.
- Corpus of Contemporary American English (2022, May 2). <https://www.english-corpora.org/coca/>
- Davies, A., Brown, A., & Elder, C. (1999). *Dictionary of Language Testing*. Cambridge University Press.

- Fan, J. (2014). Chinese test takers' attitudes towards the Versant English Test: a mixed-methods approach. *Language Testing in Asia*, 4(1), 1-17.
- Hughes, A. (1989). Testing for language teachers. Cambridge University Press.
- Hughes, A. (2003). Testing for language teachers (2nd ed.). Cambridge University Press.
- Imsa-ard, P. (2020). Voices from Thai EFL Teachers: Perceptions and Beliefs towards the English Test in the National Examination in Thailand. *LEARN Journal*, 13(2), 269-287.
- International Language Testing Association. (2021, June 29). ILTA Code of Ethics. <https://www.iltaonline.com/page/CodeofEthics>
- Lunrasri, Y. (2014). Washback effects of the Ordinary National Educational Test (O-NET) on English language teaching and learning in ninth grade (Master's dissertation, Chulalongkorn University, Thailand). Retrieved from <http://cuir.car.chula.ac.th/handle/123456789/44633>
- Mosier, C. L. (1947). A critical examination of the concept of face validity. *Educational and Psychological Measurement*, 7, 191-206.
- National Institute of Education Testing Service. (2020, September 3). What is GAT/PAT? What is on the exam?. <https://www.niets.or.th/th/content/view/1331> [In Thai]
- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22(4), 289-293.
- Oxford Learner's Word Lists. (2022, May 2). Oxford 3000 and 5000 The most important words to learn in English. <https://www.oxfordlearnersdictionaries.com/wordlists/oxford3000-5000>
- Prapphal, K. (2008). Issues and trends in language testing and assessment in Thailand. *Language testing*, 25(1), 127-143.
- Sato, T., & Ikeda, N. (2015). Test-taker perception of what test items measure: a potential impact of face validity on student learning. *Language Testing in Asia*, 5(1), 10.
- Shohamy, E. (1993). The Power of Tests: The Impact of Language Tests on Teaching and Learning. NFLC Occasional Papers.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341.
- Vongpumivitch, V. (2012). Motivating Lifelong Learning of English? Test Takers' Perceptions of the Success of the General English Proficiency Test. *Language Assessment Quarterly*, 9(1), 26-59.

- Watanabe, Y. (2004). Teacher factors mediating washback. *Washback in language testing: Research contexts and methods*, 129-146.
- Xie, Q. (2011). Is test taker perception of assessment related to construct validity?. *International Journal of Testing*, 11(4), 324-348.
- Yan, X., Thirakunkovit, S. P., Kauper, N. L., & Ginther, A. (2016). What do test-takers say? Test-taker feedback as input for quality management of a local oral English proficiency test. In *Post-admission language assessment of university students* (pp. 113-136). Springer, Cham.
- Zhou, Y., Yoshitomi, A. (2019). Test-taker perception of and test performance on computer-delivered speaking tests: the mediational role of test-taking motivation. *Language Testing in Asia*, 9(10).

Appendices

Appendix A

Questionnaire for student

แบบสอบถามความรู้/ความสามารถที่ต้องใช้ในการทำข้อสอบ

General Aptitude Test (GAT) ส่วนที่ 2: ภาษาอังกฤษเพื่อการสื่อสาร

เรียนผู้เข้าร่วมวิจัย

แบบสอบถามนี้เป็นส่วนหนึ่งของงานวิจัยตามหลักสูตรศิลปศาสตรมหาบัณฑิต สาขาวิชาภาษาศาสตร์ประยุกต์ มหาวิทยาลัยมหิดล มีวัตถุประสงค์เพื่อศึกษาว่านักเรียนและครูผู้สอนภาษาอังกฤษมีความคิดเห็นเกี่ยวกับความรู้ความสามารถที่ต้องใช้ในการทำข้อสอบวิชาภาษาอังกฤษตรงกันมากน้อยเพียงใด และเพื่อศึกษาว่าเนื้อหาข้อสอบลักษณะใดก่อให้เกิดความเข้าใจไม่ตรงกัน นอกจากนี้ยังศึกษาว่าคะแนนสอบมีความสัมพันธ์กับความคิดเห็นของผู้ทำข้อสอบอย่างมีนัยสำคัญหรือไม่ ซึ่งข้อมูลของผู้เข้าร่วมวิจัยจะนำไปใช้เพื่อประโยชน์ทางวิชาการในงานวิจัยนี้เท่านั้น และจะไม่ถูกเผยแพร่แต่อย่างใด ทางผู้ทำวิจัยใคร่ขอขอบพระคุณท่านเป็นอย่างสูงที่ให้ความร่วมมือในการตอบแบบสอบถามมา ณ โอกาสนี้

โปรดอ่านคำชี้แจงให้เข้าใจก่อนตัดสินใจลงมือทำ

1. ผู้เข้าร่วมวิจัยสามารถตัดสินใจหยุด/ยกเลิกการทำแบบสอบถามนี้ได้ทุกเมื่อ
2. คะแนนจากการทำข้อสอบฉบับนี้จะไม่ส่งผลใดๆต่อวิชาภาษาอังกฤษที่ผู้เข้าร่วมวิจัยกำลังเรียนอยู่
3. แบบสอบถามเล่มนี้ประกอบด้วย 2 ส่วน ดังนี้
 - I. ข้อมูลทั่วไปของผู้เข้าร่วมวิจัย
 - II. กระดาษคำตอบ
4. หากท่านตัดสินใจตอบและส่งแบบสอบถามนี้แล้ว จะถือว่าท่านสมัครใจยินยอมเข้าร่วมการวิจัยนี้
5. หากมีข้อสงสัยใดๆ สามารถโทรศัพท์เพื่อสอบถามผู้ทำวิจัย (นาง ศศิธร แจกสัน) ที่เบอร์ 061-410-9932

ส่วนที่ 1 ข้อมูลทั่วไปของผู้เข้าร่วมวิจัย

ชื่อ _____ นามสกุล _____

เบอร์ติดต่อ _____ Line ID _____

คณะ _____ วิชาเอก _____

คะแนน GAT ภาษาอังกฤษ _____ (ถ้าจำไม่ได้ โปรดประมาณ)

ท่านยินดีให้ผู้ทำวิจัยติดต่อท่านเพื่อขอสัมภาษณ์เพิ่มเติมเกี่ยวกับแบบสอบถามนี้ ☐ ยินดี ☐ ไม่สะดวก

ส่วนที่ 2 ข้อสอบและแบบสอบถาม

คำสั่ง

- ข้อสอบนี้มีจำนวน 17 ข้อ ท่านจงตอบคำถามแต่ละข้อโดยวงกลมล้อมรอบคำตอบข้อ a, b, c, หรือ d ตามที่คิดว่าเป็นคำตอบที่ถูกต้อง
- เมื่อตอบเสร็จแล้วให้ท่านแสดงความคิดเห็นเกี่ยวกับข้อสอบแต่ละข้อโดยใช้เครื่องหมาย / ในช่องที่มีข้อความ เห็นด้วยอย่างยิ่ง, เห็นด้วย, ไม่แน่ใจ, ไม่เห็นด้วย, หรือ ไม่เห็นด้วยอย่างยิ่ง
- กรณีตอบว่าไม่เห็นด้วย หรือไม่เห็นด้วยอย่างยิ่ง ให้ระบุว่าท่านคิดว่าข้อสอบข้อนั้นๆวัดความรู้เรื่องอะไร

ตัวอย่างการตอบข้อสอบและแบบสอบถาม

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรด ระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
1	A: _____? B: She's lovely. I really like her. a. How do you see Jane b. How do you meet Jane c. How do you go with Jane (d) How do you find Jane	ความรู้เรื่อง สำนวน และ ความหมาย ที่เหมาะสม ในบริบท				/		ความหมาย ต่างๆของ find
2	Choose the alternative which has the same meaning as the underlined word in the given sentence.	ความรู้เรื่อง คำศัพท์	/					

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรด ระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
	A New York police officer is armed with a gun and a <u>club</u>. a. I always work out at that sports <u>club</u>. b. He was hit with a heavy <u>club</u>. c. His new golf <u>club</u> is very expensive. d. The next card player is waiting for a <u>club</u>.	และ ความหมาย ที่เหมาะสม ในบริบท						

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
1	A: I've been very busy. So much work to finish before the exams. B: _____. You'll be surprised how quickly it'll all be over. a. So far so good b. Good for you c. Hang in there d. Get back on track	ความรู้เรื่อง คำศัพท์/ สำนวน และ ความหมาย ที่เหมาะสม ในบริบท						

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่เห็นใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
	Susan: What are you doing? Supa: I'm studying for the final exams. Susan: _____ 2 _____! I haven't done anything yet, I'm not in the mood. Supa: Well, the exams aren't that far off. _____ 3 _____ I won't have enough time to prepare. Susan: Do you think the exams will be difficult? Supa: Of course. That's why I'm preparing for them. Susan: Well then. I think it's a good idea to study for the exams now. Thanks for reminding me.							
2	a. How clever I am b. How happy you are c. What a good student d. What a nice occasion	เข้าใจเรื่อง ความ สอดคล้อง ของบท สนทนา						
3	a. I'm sorry b. I'm afraid c. I can't imagine d. I can't believe	เข้าใจความ สอดคล้องใน						

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
		ประโยคที่ โจทย์ถาม						
	Paula: Linda, did you manage to get hold of Jane after work yesterday? Linda: No, I'm sorry, Paula. <u>4</u> last night. And this morning I tried again, but nobody answered. Paula: Don't you have her cell phone number? Linda: I'm sorry, I don't. Have you got it? Paula: Of course I do. You should have called me last night. Linda: Well, I thought you were busy preparing for the meeting, and I didn't want to disturb you. <u>5</u> _____? Paula: The meeting starts in 10 minutes. That's out of the question. She'll never get here in time.							
4	a. I didn't put her through b. I didn't have her number c. Her line was always busy d. Her phone wouldn't answer	ความรู้ เกี่ยวกับ สำนวนที่ใช้						

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
		ใน ชีวิตประจำวัน และ ความหมาย ที่เหมาะสม ในบริบท						
5	a. Should I call her now b. When is the meeting c. How soon is the meeting d. Would you take her call	ความสามารถ ในการ เลือก ประโยค คำถามที่มี ความหมาย เหมาะสมใน บริบท						
6	Three of the words in each group relate to one another in their meaning. Choose the word that does NOT belong. a. grab b. seize c. snatch d. rise	ความหมาย ของคำศัพท์ โดยไม่มี บริบทช่วย						

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่เห็นใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
7	Choose the alternative which has the same meaning as the underlined word in the given sentence. <u>Apply</u> some moisturizer on your skin when it gets dry. a. Students should <u>apply</u> what they have learned in their real life. b. I had to <u>apply</u> the brake very quickly. c. I will <u>apply</u> a plaster on the cut to your finger. d. My sister wants to <u>apply</u> for a new job.	ความรู้เรื่อง คำศัพท์ และ ความหมาย ที่เหมาะสม ในบริบท						
8	Choose the best alternative to make the sentence meaningful. Valuable _____ of oil and gas lies hidden in rocks on the _____ where they are tapped by drilling down through those rocks. a. reservoirs – seabed b. collections - sea wall c. resources - sea level d. assets - seaboard	ความรู้เรื่อง คำศัพท์ และ ความหมาย ที่เหมาะสม ในบริบท						

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
9	 What can be inferred from the comic strip? - The girl doesn't say no. - The boy doesn't say no. - The boy doesn't like the girl. - The girl doesn't like the boy.	เข้าใจ ความหมาย ที่เหมาะสม ในบริบท และเข้าใจ ใจความ สอดคล้อง ของบท สนทนา						
10	 What is TRUE according to this sign? - You are free to smoke in this area. - There is no air pollution in this area. - You are not allowed to smoke in this area. - Free cigarettes are available in this area.	ความสามารถ ในการอ่าน ป้ายและ เข้าใจ ความหมาย ของสำนวน						
11-13	My family and I have watched with interest the swallows that, over the years, have sought out a nest attached to the wall underneath the small terrace of							

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
	our house. They come and go, and we look forward to their return each spring. One day I noticed that the nest had fallen over with the small occupants - six fat fledglings covered in good plumage - still inside. My eight- year-old son arranged an emergency shelter: he put the babies inside a cardboard box, along with bits of cotton wool and a few pieces of the clay from the destroyed nest. When my husband returned home that evening, we considered how we could build a new nest for the swallows. After some animated discussion and without hope of success, we decided to put the baby birds in another deeper cardboard box, together with the clay and soft cotton wool. With a few nails in the wall and a few holes in the improvised nest, the box was attached to the wall with its little occupants inside. We worried that our handling of the tiny chicks had left the odor of humans on them. Would the parents, when they returned, become							

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
	suspicious, or even reject their young ones? So we waited anxiously. With dawn came the sound of the parents busily going about the job of feeding six hungry birds! They looked after their brood until the fledglings became self-sufficient; only then did they dedicate themselves to fixing up the new nest we had built. Some 22 years later, this nest is still sought out by successive generations of swallows who come and depart- the source of renewed happiness and joy for us.							
11	What is the first paragraph mainly about? a. A family activity b. A shelter for swallows c. Environment protection d. The small terrace of a house	เข้าใจ ใจความหลัก (main idea) ของ ย่อหน้าที่ อ่าน						
12	What was NOT used to make a new nest for the baby swallows? a. A terrace b. A cardboard box	เข้าใจใน รายละเอียด						

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่เห็นใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
	c. Cotton wool d. Clay	ของย่อ หน้าที่อ่าน						
13	What can be concluded from the passage? a. Environmental protection is necessary. b. A friendship across species is possible. c. Swallows are man's best friends. d. Life is not easy for swallows.	เข้าใจเนื้อ เรื่องโดย สรุป						
14- 15	(1) A recent study found that resident doctors know when they are exhausted, upset or overwhelmed and when they are, they are far more likely to admit they made an error. What they are admitting is that sleepiness is not the only factor. Above and beyond that is general distress and mental fatigue, researchers reported in the Journal of the <i>American Medical Association</i> . (2) “While fatigue is important, there is this whole domain of distress beyond fatigue that also demands attention,” Colin West of the Mayo Clinic in Rochester, Minnesota, who led the study, said in an interview. The findings may help point to ways to better reduce							

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
	the burdens on resident doctors, known in some countries as junior doctors, and in turn prevent mistakes. “I think this is going to have an impact on healthcare reform,” Dr. West said. “We need to put resources into training and medicine to control work hours and maintain physician well-being.” (3) The US Institute of Medicine reported in 1999 that between 48,000 and 98,000 Americans die each year from preventable medical errors ranging from drug overdoses to infections caught in the hospital. And doctors, unions and other experts have been clamoring to cut the work hours of residents, who are required to work at least 50 hours a week at most training hospitals. (4) Dr. West and colleagues surveyed 356 residents at 163 medical schools globally. They found that 39% reported making at least one major medical error during the study period, and these doctors were also more likely to say they were sleepy, fatigued, or							

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
	stressed “What we have shown in this data is that fatigue is important, but it's only part of the issue and previous studies have not paid much attention to the distress factors,” Dr West said “Over the course of their training and the course of their career everyone has made a major error. Everyone tries to be perfect but no one is. He said only limiting work hours may not be enough. (5) Teaching hospitals across the United States have moved to limit residents’ work weeks to 40 hours to reduce fatigue-related errors. It is estimated that this could cost US\$171 million to \$487 million a year, depending on whether additional residents are hired or substitute providers are brought in.							
14	Which word is closest in meaning to "clamoring" (paragraph 3)? a. Demanding b. Rejecting c. Announcing d. Encouraging	ความรู้เรื่อง คำศัพท์และ ความหมาย ที่เหมาะสม ในบริบท						

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
15	What does "this" (paragraph 5) refer to? a. Trying to be perfect b. Being sleepy, fatigued, or stressed c. Limiting residents' work hours d. Hiring additional residents	เข้าใจ รายละเอียด ของเนื้อหา และเข้าใจ ความสัมพันธ์ ของ ข้อความ						
16	Read the following statements and choose the underlined part that is grammatically wrong. If buildings (a) <u>sprang up</u> suddenly out of the ground (b) <u>like</u> mushrooms, their rooftops would be covered (c) <u>into</u> a layer of (d) <u>soil</u> and plants.	ความรู้เรื่อง โครงสร้าง ประโยคและ ไวยากรณ์ ภาษาอังกฤษ ฯ						
17	Choose the best answer to complete the sentence. The government spent much more in the last quarter than planned _____ it spent considerably less in the previous one. a. whereas b. before c. until d. when	เข้าใจ ความหมาย และ ความสัมพันธ์ ของ ข้อความเพื่อ						

ข้อที่	เนื้อหาข้อสอบ	ประเด็นที่ คำถามข้อนี้ ต้องการ ทดสอบคือ	เห็นด้วยอย่างยิ่ง	เห็นด้วย	ไม่แน่ใจ	ไม่เห็นด้วย	ไม่เห็นด้วยอย่างยิ่ง	โปรดระบุ ความรู้ เรื่องอื่นที่ จำเป็นใน การตอบ คำถาม
		เลือกใช้ คำเชื่อมได้ ถูกต้อง						

ส่วนที่ 3 ความคิดเห็นทั่วไปเกี่ยวกับข้อสอบ

คำสั่ง

ให้ท่านแสดงความคิดเห็นเกี่ยวกับข้อสอบโดยใช้เครื่องหมาย / ในช่องที่มีข้อความ มากที่สุด, มาก, ไม่แน่ใจ, น้อย, หรือ น้อยที่สุด

ข้อที่	ความคิดเห็น	มากที่สุด	มาก	ไม่แน่ใจ	น้อย	น้อยที่สุด
1	คุณคิดว่าข้อสอบฉบับนี้ทดสอบความรู้ความสามารถในการเขียนเป็นภาษาอังกฤษของคุณในระดับมากน้อยเพียงใด					
2	คุณคิดว่าข้อสอบฉบับนี้ทดสอบความรู้ความสามารถในการพูดภาษาอังกฤษของคุณในระดับมากน้อยเพียงใด					
3	คุณคิดว่าข้อสอบฉบับนี้ทดสอบความรู้ความสามารถในการสื่อสารภาษาอังกฤษของคุณในระดับมากน้อยเพียงใด					

โปรดระบุ ความรู้ความสามารถในการสื่อสารภาษาอังกฤษด้านที่ท่านคิดว่าข้อสอบฉบับนี้ต้องการทดสอบ

Appendix B

Questionnaire for teachers

แบบสอบถามความรู้/ความสามารถที่ต้องใช้ในการทำข้อสอบ

General Aptitude Test (GAT) ส่วนที่ 2: ภาษาอังกฤษเพื่อการสื่อสาร

เรียนผู้เข้าร่วมวิจัย

แบบสอบถามนี้เป็นส่วนหนึ่งของงานวิจัยตามหลักสูตรศิลปศาสตรมหาบัณฑิต สาขาวิชาภาษาศาสตร์ประยุกต์ มหาวิทยาลัยมหิดล มีวัตถุประสงค์เพื่อศึกษาว่านักเรียนและครูผู้สอนภาษาอังกฤษมีความคิดเห็นเกี่ยวกับความรู้ความสามารถที่ต้องใช้ในการทำข้อสอบวิชาภาษาอังกฤษตรงกันมากน้อยเพียงใด และเพื่อศึกษาว่าเนื้อหาข้อสอบลักษณะใดก่อให้เกิดความเข้าใจไม่ตรงกัน นอกจากนี้ยังศึกษาว่าคะแนนสอบมีความสัมพันธ์กับความคิดเห็นของผู้ทำข้อสอบอย่างมีนัยสำคัญหรือไม่ ซึ่งข้อมูลของผู้เข้าร่วมวิจัยจะนำไปใช้เพื่อประโยชน์ทางวิชาการในงานวิจัยนี้เท่านั้น และจะไม่ถูกเผยแพร่แต่อย่างใด ทางผู้ทำวิจัยใคร่ขอขอบพระคุณท่านเป็นอย่างสูงที่ให้ความร่วมมือในการตอบแบบสอบถามมา ณ โอกาสนี้

โปรดอ่านคำชี้แจงให้เข้าใจก่อนตัดสินใจลงมือทำ

1. ผู้เข้าร่วมวิจัยสามารถตัดสินใจหยุด/ยกเลิกการทำแบบสอบถามนี้ได้ทุกเมื่อ
2. คะแนนจากการทำข้อสอบฉบับนี้จะไม่มีผลใดๆต่อการสอนภาษาอังกฤษที่ผู้เข้าร่วมวิจัยกำลังสอนอยู่
3. แบบสอบถามเล่มนี้ประกอบด้วย 2 ส่วน ดังนี้
 - I. ข้อมูลทั่วไปของผู้เข้าร่วมวิจัย
 - II. กระจายคำตอบ
4. หากท่านตัดสินใจตอบและส่งแบบสอบถามนี้แล้ว จะถือว่าท่านสมัครใจยินยอมเข้าร่วมการวิจัยนี้
5. หากมีข้อสงสัยใดๆ สามารถโทรศัพท์เพื่อสอบถามผู้ทำวิจัย (นาง ศศิธร แจกสัน) ที่เบอร์ 061-410-9932

ส่วนที่ 1 ข้อมูลทั่วไปของผู้เข้าร่วมวิจัย

ชื่อ _____ นามสกุล _____

เบอร์ติดต่อ _____ Line ID _____

กำลังสอนอยู่ที่ โรงเรียน/มหาวิทยาลัย _____

มีประสบการณ์ในการสอนภาษาอังกฤษมาเป็นเวลา _____ ปี

วิชาภาษาอังกฤษที่เคยสอนหรือกำลังสอนอยู่ในปัจจุบัน _____

ท่านยินดีให้ผู้ทำวิจัยติดต่อท่านเพื่อขอสัมภาษณ์เพิ่มเติมเกี่ยวกับแบบสอบถามนี้ ยินดี ไม่สะดวก

ส่วนที่ 2 ข้อสอบและแบบสอบถาม

This part is as same as the students' questionnaire part II

ส่วนที่ 3 ความคิดเห็นทั่วไปเกี่ยวกับข้อสอบ

คำสั่ง

ให้ท่านแสดงความคิดเห็นเกี่ยวกับข้อสอบโดยใส่เครื่องหมาย / ในช่องที่มีข้อความ มากที่สุด, มาก, ไม่แน่ใจ, น้อย, หรือ น้อยที่สุด

ข้อที่	ความคิดเห็น	มากที่สุด	มาก	ไม่แน่ใจ	น้อย	น้อยที่สุด
1	คุณคิดว่าข้อสอบฉบับนี้ทดสอบความรู้ความสามารถในการเขียนเป็นภาษาอังกฤษของนักเรียนในระดับอย่างน้อยเพียงใด					
2	คุณคิดว่าข้อสอบฉบับนี้ทดสอบความรู้ความสามารถในการพูดภาษาอังกฤษของนักเรียนในระดับอย่างน้อยเพียงใด					
3	คุณคิดว่าข้อสอบฉบับนี้ทดสอบความรู้ความสามารถในการสื่อสารภาษาอังกฤษของนักเรียนในระดับอย่างน้อยเพียงใด					

Appendix C

Interview Questions

1. You were not sure/disagreed/strongly disagreed with the stated ability being tested in item number ... Can you explain why you chose that?
2. Could you explain why you gave such rating for that this test could measure writing/speaking/communication abilities?

The additional question for high-school and university teachers

1. How does the test have an impact on your teaching/your assessment?

The additional question for students

1. How does the test have an impact on your life/your study life?