

## Students' Perceptions of Traditional English Language Testing in Thailand

Natjiree Jaturapitakkul, Ph.D.

School of Liberal Arts, King Mongkut's University of Technology Thonburi,  
Bangkok, Thailand  
E-mail: natjiree.jat@kmutt.ac.th

Doi:10.5901/ajis.2013.v2n3p445

### Abstract

*This study investigates students' perceptions towards traditional English language testing particularly in Thailand. A survey of 323 randomly selected Thai freshmen from various programmes was conducted. An open-ended questionnaire with four main concerns—test fairness, test format, test content validity, and learning opportunities from the tests—was used to collect data. All responses were transcribed, analysed for frequency and categorised into salient themes. The results showed that current testing practices were viewed to be fair enough as a good way to measure students' English ability. The formats of tests were generally perceived as a reflection of their English ability though there should be more various and new forms of item types suitable for all skills tested. Furthermore, current tests were viewed to be capable of measuring important content students learnt in courses. Regarding whether existing tests promoted learning, there were opportunities to learn from the tests. However, students disliked tests of memorisation of knowledge, preferring tests that measured cognitive abilities. The results suggest practical guidelines to improve existing testing practices and even to inspire us to create innovative kinds of tests.*

### 1. Background

It is inevitable that teaching involves assessment (especially in the form of exams and tests). We are now in a world where many young people in many countries go through school and higher education and are faced with an unprecedented number of exams and tests. In fact, this is a world which has been developed into a testing society (Broadfoot, 2005). Testing in education is an attempt to measure a person's knowledge, intelligence, or other characteristics in a systematic way. Testing is extremely vital in education since it exerts a greatly powerful influence in societies (McNamara and Roever, 2006), acts as an influential implicit policy instrument (Darasawang and Watson Todd, 2012), plays a major role in the implementation and introduction of language policies and practice (Paran, 2010), and often utterly affects language learning and teaching in classrooms. The impacts of testing on educational systems can also be seen in English language education in the countries of South-East Asia like Thailand.

In Thailand's present situation where traditional testing exists, it is unavoidable for students to take a test as a requirement of their education at the end of a predetermined period of instruction (for example, mid-term and final). Traditional testing takes the form of paper-and-pencil tests which are typically used for the assessment of separate components of language (grammar, vocabulary) and receptive understanding (listening & reading comprehension). Test items in such tests or even in standardized tests are often in fixed response format or one-answer questions (e.g. multiple choice and true/false). Students are rated in relation to how many right answers are given.

Even though traditional tests have a long history and have proven very useful to some extent, they receive much attention from educators and policymakers. Criticism has been directed both at the limitations of the tests and at their influence on students' learning. For example, misjudging a student's ability from a one-shot test, revealing only whether the student can recognize and recall what was learned out of context, fostering only extrinsic motivation, limiting to paper-and-pencil, one-answer questions and being more like drills, assessing static and too-often arbitrarily discrete points (Wiggins, 1990), and lacking assessment of productive skills and ignoring the assessment of higher-order thinking (Jaturapitakkul and Watson Todd, 2012). Furthermore, in spite of the ubiquity of testing, there is nevertheless "a widespread perception that the needs of teachers and learners are not currently well served by assessment practice and by assessment research" (McNamara, 2001: 340). It is then very appealing to further investigate what is wrong with current testing practices and how they could serve the needs of both teachers and learners.

Most previous research studies and published articles in accordance with traditional assessment have been derived from the views of researchers as well as educators. In this present study, however, the views of students as a

large important group of stakeholders who have ordinarily been disregarded (Pino-Silva, 2008) were investigated regarding their preferences and problems related to current English language testing. The results might shed some lights in English language testing and could primarily inform ways to improve current testing practices. Implications for testing, teaching and learning would also be derived from this study.

## 2. Literature Review

The review of literature draws on theoretical foundations of this study which include the concepts of test fairness, test format, test content validity and learning validity. The fairness of the test literature adds to this study by providing an insight into how the tests are viewed as a fair way of measuring students' English ability. The format of the tests gives an idea about what item test type has been used in the tests in order to help reflect English ability in all skills. Key principles underpinning test content validity also canvass an idea for measuring what students learnt in class. The learning validity provides an insight into opportunities to learn from tests and how test takers can learn through the tests. Each of these four aspects will be considered in turn and collectively used as a framework for designing a research instrument.

### 2.1 Test fairness

The test fairness has been defined in many different ways. It is broadly conceived in the Standards for Educational and Psychological Testing as freedom from any kind of bias and equitable treatment of all test takers (regardless of race, religion, gender, or age) in the testing process for all stages (Xi, 2010). Fairness can also refer to whether test scores measure the same thing across different groups. This is what psychometricians called "construct validity". However, it can also be applied at the individual level as Penfield and Camilli (2007) suggested. For example, a test taker should be given a fair chance to understand the test questions, but testing conditions or language issues might pose obstacles to this goal.

### 2.2 Test format

In general, traditional English language tests are in a form of fixed response format with one correct answer. In other words, it is a kind of objective item. Objective items require students to select the correct response from several alternatives or to supply a word or short phrase to answer a question or complete a statement (Ory, 2004). Examples of objective items include multiple-choice, true-false, matching and completion. These item test types can help assess students' English ability in all skills to a certain extent and can contribute to ease of scoring with more objectivity. In most tests (even in standardised tests) objective items are more dominant than subjective ones. When designing achievement tests, test content should be based on the objectives of the course.

### 2.3 Test content validity

One of the core characteristics of a test that needs to be of concern is its content. The consideration of the test content is an important part of both test development and test use. According to Bachman (1990), there are two aspects that should be included to ensure that the content is valid: content relevance and content coverage. In other words, the test should be relevant to and cover a given area of content or ability that purports to be measured. This type of validity is mostly concerned with achievement tests that directly relate to the content taught in courses. It is used specifically when a test is attempting to measure a defined domain or teaching/learning objectives in a course and indicates whether or not the test items adequately represent that domain or course objectives.

### 2.4 Learning validity

The aforementioned key criticism of traditional testing is its negative washback effects. To some extent, this is due to disjunction between testing and learning. Traditionally, assessment in the form of exams or tests is viewed as an aspect separate from learning and does not provide opportunities for test-takers to learn from the assessment. If assessment is to truly aid learning, there needs to be a move away from this traditional model of assessment of learning to a new model of assessment for learning (Paran, 2010). In other words, in addition to measuring previous learning, tests should provide opportunities for test-takers to learn through the test. This has been termed the learning validity of a test (Tomlinson,

2005), and such assessment-driven learning may occur while preparing for the test, while taking the test, or on receiving feedback from a test.

### 3. Research Questions

This research used a survey technique to investigate students' views on traditional English language testing in Thailand particularly in four main areas: test fairness, test format, test content validity and opportunities to learn from tests. Incorporated with the purpose of the study, four research questions were formulated.

1. What are students' views on fairness in traditional English language tests?
2. What are students' views on the format used in traditional English language tests?
3. What are students' views on content validity in traditional English language tests?
4. What are students' views on learning opportunities from traditional English language tests?

### 4. Research Methodology

#### 4.1 Participants

Three hundred twenty-three first-year undergraduate students at a Thai university in Bangkok were randomly selected to participate in this study. These students took the second English compulsory course during the 2012 academic year. They were approximately the same age (17-18 years old) with mixed language ability and from faculties of Engineering and Science which are two of the major fields in Thai universities. All of them use English as a foreign language and have been exposed to traditional English language tests for more than ten years.

#### 4.2 Instrument

To avoid constraining students' responses, an open-ended questionnaire was constructed to investigate the students' perceptions towards traditional English testing in Thailand. The questions generally focussed on four key issues in English language testing: test fairness, test format, test content validity, and learning opportunities from the tests. The questions included in the questionnaire are:

1. What do you think about the fairness of the tests you have taken? In other words, are they a good way of measuring your English ability? How could they be improved to increase fairness?
2. What do you think about the format of the tests you have taken i.e. multiple choice, T/F and matching? Do they really help reflect your actual English ability in all skills? How could they be improved to test your English ability?
3. Can tests measure important content you learnt in English courses? In other words, are they a good way of measuring what you learnt? Why?
4. Do tests help you to learn? How could they be improved to promote learning?

The content of the questionnaire was validated by three experts. It was then revised, translated into Thai, and piloted with ten students before a main study was done. This was to reassure the validity of the instrument and the questionnaire's intelligibility.

#### 4.3 Data Collection

The survey was conducted in the form of a paper-based questionnaire from March-April, 2012. The questionnaires were administered in regular class hours. The purpose of the survey was described to the random samples and the questions clarified. They were also informed that their questionnaire responses would be confidential and would not affect their English grades. The questionnaire took about 15 minutes to complete.

#### 4.4 Data Analysis

All responses were transcribed and saved in Microsoft Word files. To facilitate data analysis, for each question, table formats were designed with columns for 1) the subject's number, 2) YES/No answers, 3) the comments offered by each subject, 4) how to improve the tests, and 5) coding. Each response was read and categorised into salient themes. Those

themes identified were coded in the formats. This made it possible to establish the frequency of theme occurrence. Nevertheless, some responses were discarded from the analysis in case of irrelevance or ambiguity.

## 5. Results and Discussion

The following presents the findings of the study according to the four research questions. Due to space limitations and a large number of various responses, only the categories with a frequency greater than 5 percent are presented and the three highest percentages are discussed. Please note that percentages were calculated from valid responses only.

### 5.1 Research Question 1: What are students' views on fairness in traditional English language tests?

The first research question to report on concerns students' views on fairness in traditional English language tests. Table 1 illustrates the frequency of perceived fairness of the tests, the categories of the responses given, the frequency of the responses, and the percentage of the total of responses. This table format will be applied to answer all research questions.

**Table 1:** Students' views on fairness in traditional English language tests

Questions	YES	%	NO	%
Are the tests you have taken a good way of measuring your English ability?	233	84.73	42	15.27
<b>Category of Responses</b> (How could they be improved to increase fairness?)	<b>Total</b>	<b>%</b>		
1. Students should be tested according to what they have learnt in English courses.	27	29.67		
2. Instructions and test questions should be clear and translated into Thai to ease understanding.	16	17.58		
3. The level of test difficulty should match with the level of students' English proficiency/ability.	12	13.19		
4. Tests should cover content for all fields of study and used across the country.	9	9.89		
5. The same set of tests (shuffled items for a different version is acceptable) should be used for the same standard across the country.	9	9.89		
6. Tests should be designed for a particular group of students/a particular program i.e. regular or international	7	7.69		
7. Answer key should be provided after the test so that they can know which items they had answered correctly.	6	6.60		
8. Test content should relate to English in daily life, world knowledge or general issues.	5	5.49		
<b>Total</b>	91	100		

From Table 1, it can be seen that the majority of respondents (84.73%) perceived that the tests they have taken are a good way of measuring their English ability. This might be because students have been exposed to and familiar with this form of traditional English tests since they entered into an educational system. The ubiquity of tests can ultimately dictate to their perceptions to automatically involve testing in all stages of their education and that everyone must be involved in this educational system. In addition, the way the tests are administered by using the same set of tests across different groups would contribute to an equitable treatment of all test takers. This would make them feel that the tests are fair enough to be valid and implied as a good way of measuring their English ability.

Regarding how to increase test fairness, the most frequently reported comment is to certainly test what students have learnt in English courses. In other words, whether or not the content of tests covers what have been taught in class. This is in accordance with what Felder (2002) and Vos (2000, cited in Bouville, 2008) commented that students deem an exam unfair when there are questions on content which is not covered in lectures or in courses. In the light of evidence from this study, students were likely to view test fairness in relation to content relevance and content coverage as it seems to be tangible to be tested according to course objective domains. The ideas of fairness of the test and test content validity seem to overlap to some extent.

Secondly, it was found that test rubrics and instructions should be clear and bilingual (in both Thai and English) to avoid misunderstandings. In other words, it might be unfair if despite the test takers knowing well the test content, they nevertheless misinterpret the instructions and the test questions. By providing Thai translations or making it as a bilingual test, this would not mislead the test-takers on the points to be asked and give a fair chance to a test taker at the individual level (Penfield and Camilli, 2007) to understand what to do or answer in the tests. This, therefore, results in the testing of actual knowledge of the learners, and not the testing of the ability to comprehend or interpret test rubrics and instructions.

Thirdly, the level of test difficulty in accordance with the level of students' English ability was reported. This comment supports an idea of Henning (1987) that 'a test should never be selected or developed without due consideration of the abilities and other characteristics of the intended examinees'. By doing so, all test takers would be given a fair chance to be tested with an appropriate test to reflect the test takers' actual ability.

5.2 Research Question 2: What are students' views on the format used in traditional English language tests?

Table 2 presents the students' views on the format used in traditional English language tests and on how the tests can be improved to test their English ability in all skills.

**Table 2:** Students' views on the format used in traditional English language tests

Questions	YES	%	NO	%
Do the test formats really help reflect your actual English ability in all skills?	250	83.61	49	16.39
<b>Category of Responses</b> (How could they be improved to test your English ability?)	<b>Total</b>		<b>%</b>	
1. More varied item types like gap-fill, short answer, or matching should be included in the tests.	54	62.07		
2. More suitable ways of testing productive skills i.e. writing and speaking	12	13.79		
3. New forms of tests should be developed (either paper-based or computer-based)	9	10.34		
4. More distractors are added into the test items.	7	8.05		
5. Item types like true/false should be avoided because they are easy to guess (50% chance of a correct answer)	5	5.75		
<b>Total</b>	87	100		

From Table 2, it is obvious that most subjects (83.61%) viewed the formats in the tests helpful to reflect their actual English ability. This might be due to students' familiarity of the test format used and their repeated encounter with the same formats, especially, the dominance of multiple choice in all skills and in all tests, at least in Thailand.

In regard to how to improve the formats in the tests, the three highest percentages reported will be discussed jointly since they have two identical denominators: variety and suitability of the test formats. Even though existing test formats were perceived as likely to reflect students' English ability, many of them were well-aware that some formats of tests could help facilitate test taking in some skills, such as multiple-choice items in reading tests, but these were not applicable in speaking, listening and writing (Jaturapitakkul and Watson Todd, 2012). In addition, in Thailand where a multiple choice format has been long dominant in all skills due to practical issues, this would not be appropriate for assessment of productive skills and higher-order thinking. As a result, looking for more variety of test formats like gap-fill, short answer or matching as well as new forms of test format (regardless of being paper- and computer-based) would be plausible. By providing more variety of item test types (with a recognition of their advantages, limitations and when to use appropriately, more details in Ory, 2004) and perhaps new test formats, this could enhance more efficient use of tests with more valid test results and be more appropriate for the objectives to be tested for each skill.

5.3 Research Question 3: What are students' views on content validity in traditional English language tests?

Table 3 reveals the students' views on content validity in traditional English language tests and reasons why they say so.

**Table 3:** Students' views on content validity in traditional English language tests

Questions	YES	%	NO	%
Can tests measure important content you learnt in English courses?	275	91.06	27	8.94
<b>Category of Responses</b> (Why are they a good way of measuring what you learnt?)	<b>Total</b>		<b>%</b>	
1. The tests students have taken were mostly based on what they learned in English courses.	90	62.07		
2. Tests are seen as a key instrument to assess how much students learnt from English courses regardless of differences of learners' English proficiency.	47	32.41		
3. Tests can help in self-diagnosis i.e. if they learnt from the courses, and what area they should improve.	8	5.52		
<b>Total</b>	145	100		

As seen in Table 3, the majority of respondents (91.06%) viewed the content in existing tests valid. In other words, to some extent the tests can measure important content in English courses they learnt in class. This corresponds to the reason given that most of the tests students had taken were dependent on content taught or cover the stated objectives in courses so that they could apply knowledge gained to tackle questions in the tests. Therefore, content validity is overwhelmingly supported in current tests in Thailand. In addition, it is even more interesting when students viewed the tests as a key instrument to assess their learning as well as diagnose their weak points in English. By relying on test content which is believed to be valid (covers the stated objectives), this would help assess their knowledge and make them realize (either during or after test taking) what they can do or cannot do (which also indicate their weaknesses) from the questions they are not capable of answering, particularly linguistic knowledge (i.e. grammar and vocabulary).

5.4 Research Question 4: What are students' views on learning opportunities from traditional English language tests?

Table 4 presents the students' views on opportunities to learn from traditional English language tests and their comments on how the tests can be improved to promote learning.

**Table 4:** Students' views on learning opportunities from traditional English language tests

Questions	YES	%	NO	%
Do tests help you to learn?	275	89.00	34	11.00
<b>Category of Responses</b> (How could they be improved to promote learning?)	<b>Total</b>	<b>%</b>		
1. Tests of memorization should be avoided, preferring tests that measure cognitive abilities or language use	26	24.07		
2. Answer key should be provided with explanation so that they can learn from it.	18	16.68		
3. Test content should be up-to-date and interesting so that they can learn from reading texts.	15	13.88		
4. All skills should be tested with a variety of item test types and cover all fields of study.	13	12.04		
5. The difficulty level of tests should be suitable for the level of students.	11	10.19		
6. Bilingual tests to ease their understanding and then they can learn from the tests.	11	10.19		
7. Test items should be mixed for both objective and subjective types.	8	7.40		
8. More computer-based tests should be incorporated.	6	5.55		
<b>Total</b>	<b>108</b>	<b>100</b>		

In Table 4, most respondents (89%) stated that there were learning opportunities from test taking. Even though there are a number of critiques in current traditional testing practices (O'Sullivan, 2011 and Paran, 2010), learning opportunities from tests can take place somehow and to some extent. Students might view opportunities to learn from tests in many ways. For example, learning from taking tests which require cognitive abilities or high-order thinking skills, learning from test items they had answered incorrectly, and learning from reading texts/passages in the tests.

In regard to ways to promote learning in tests, avoiding tests of memorization and instead preferring tests that measure cognitive abilities or language use in productive skills was proposed as the highest percentage. As seen in many contexts in Thailand, most tests are dominated by memorization resulting in the wide use of multiple choice format in all educational levels regardless of the insignificant inclusion of other formats in tests. This phenomenon promotes ephemeral memorisation of surface knowledge as Jaturapitakkul and Watson Todd (2012) referred to and would limit language objectives assessed to only reading and linguistic knowledge. Therefore, there should be a move away from testing memorisation to testing cognitive or high-order thinking abilities.

In the second place, an answer key should be provided with explanation so that this could enhance learning opportunities from tests. This relates to unclear current marking practices in Thailand. To clarify, in traditional testing particularly in mid-term and final examinations, students have never been told what the right answers in the tests are. When they do not know which items they had answered correctly or incorrectly (preferably with explanation), there is no learning taking place.

The third highest percentage reported was to incorporate updated and interesting texts in tests so that students can learn from the content in the texts. For this comment, some students viewed learning not only in forms of linguistic knowledge which is assessed in the tests but also from reading texts in the tests. This might broaden their vision and perspectives towards our world whose information has been changing and updating constantly.

## 6. Implications

From the findings of this study, this could bring us to some implications and recommendations for both pedagogical and testing aspects as follows:

- Teachers must strictly be concerned with course objectives from the very beginning. They should teach and test in relation to course objectives or what purports to be measured.
- When designing the tests, it should be assured that the test content is relevant to what is taught in class and covers the objectives that aim to be assessed and can be testable by means of tests. More integration of various item formats and more appropriate ways to test productive skills would be preferable.
- To increase more opportunities of assessing cognitive abilities as well as productive skills, some test formats like interview and essay can be arranged in the form of continuous assessment or formative assessment which can happen during and continue throughout a semester.
- There should be a move from paper-based to computer-based testing. Some advantages of computer-based tests might compensate for the missing values of paper-based tests. For example, providing immediate scores with explanation, or giving hints or help such as glossary to facilitate test taking. Doing so can promote learning from tests.
- Some more research studies regarding perceptions of traditional English language testing should be conducted with more subjects in this field in order to gain more responses and new target groups like secondary students. This might present more valid data and yield different findings. Moreover, other aspects of English language testing like reliability, washback, authenticity and practicality can be investigated to gain more insights and shed light on testing practices in Thailand.

## 7. Conclusion

This study aimed to investigate students' perceptions of current testing practices in Thailand. It was found that existing tests were viewed to be fair as a measurement of students' English ability. The formats of tests were generally perceived as a reflection of their English ability, however, it was preferred to incorporate more varied test formats in order to test all skills. Furthermore, the content of the tests were viewed to be valid—testing what they had learned in English courses. Lastly, there were opportunities to learn from the tests to some extent. The results suggest practical guidelines to improve existing testing practices.

## 8. Acknowledgement

The author would like to express her gratitude for the funding to the School of Liberal Arts, King Mongkut's University of Technology Thonburi (Project no. 2555102).

## References

- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bouville, M. (2008). Exam fairness. Retrieved September 11, 2013, from <http://arxiv.org/pdf/0803.4235.pdf>
- Broadfoot, P.M. (2005). Dark alleys and blind bends: Testing the language of learning. *Language Testing* 22(2), 123-145.
- Darasawang, P. and Watson Todd, R. (2012). The effect of policy on English language teaching at secondary schools in Thailand. In Low, E.-L. and Hashim, A. (eds.) *English in Southeast Asia: Features, Policy and Language in Use* (pp. 207-220). Amsterdam: John Benjamins.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Cambridge: Newbury House Publishers.
- Jaturapitakkul, N. and Watson Todd, R. (2012). Support adaptive testing: Towards a new future in language education. *Proceedings of the 4<sup>th</sup> Language in the Realm of Social Dynamics International Conference 2012* (pp. 54-59). Bangkok.
- McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing* 18(4), 333-349.
- McNamara, T. & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell.
- O'Sullivan, B. (2011). *Language testing: Theories and practices*. Great Britain: Palgrave Macmillan.
- Ory, J. (2004). Improving your test questions. EES Memorandum #27. Evaluation and Examination Service. University of Iowa. Retrieved from [http://www.uiowa.edu/~examserv/resources\\_fees/Technical\\_Bulletins/Improving\\_Test\\_Questions.html](http://www.uiowa.edu/~examserv/resources_fees/Technical_Bulletins/Improving_Test_Questions.html)
- Paran, A. (2010). More than language: The additional faces of testing and assessment in language learning and teaching. In Paran, A. and Sercu, L. (eds.) *Testing the Untestable in Language Education* (pp. 1-13). Bristol: Multilingual Matters.
- Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In S. Sinharay & C.R. Rao (eds.) *Handbook of Statistics*

Volume 26:

Psychometrics (pp. 125-167). New York: Elsevier.

Pino-Silva, J. (2008). Student perceptions of computerized tests. *ELT Journal* 62(2), 148-156.

Tomlinson, B. (2005). Testing to learn: A personal view of language testing. *ELT Journal* 59(1), 39-46.

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research & Evaluation* 2(2). Retrieved from <http://PAREonline.net/getvn.asp?v=2&n=2>

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing* 27(2), 147-170.