

**THE IELTS READING TEST: A REVIEW ON ITS DEVELOPMENT
AND LANGUAGE TEST PERFORMANCE**

Yuliar Masna¹

Abstract

International English Language Testing System (IELTS) is regarded as one of proficiency tests that is generally accepted by British and Australian universities in selecting and screening overseas students. Changes in the tasks and texts of IELTS Reading test have been improved due to response to development in applied linguistics. In constructing the module of the IELTS reading test, there are several dimensions that should be considered by developers such as British Council; they are the validity, reliability, fairness, and impact of the test. This is very important in order that test takers can achieve more valid, reliable, and fair band scores from this section test. Therefore, this writing is aimed to review the development and language test performance of IELTS, particularly focusing on the IELTS Reading Test section.

Keywords: *proficiency test, validity, reliability*

¹ Yuliar Masna dosen FTK UIN Ar-Raniry

INTRODUCTION

International English Language Testing System (IELTS) is a proficiency test used by English-speaking countries such as the United Kingdom and Australia to recruit international students in universities. Shaw and Falvey (2008) explain a brief history of IELTS. In its development, IELTS has been revised from several times. In 1989, IELTS first became operational; it included reading and writing as specialized module, and listening and speaking as non-specialized module. In April 1995, further revisions to test were implemented to respond to developments in applied linguistics, measurement theory, and teaching practice. In 1998, revision project for speaking module is first launched and employed; then in July 2001 the revised IELTS speaking test was introduced.

IELTS is centrally organized by the British Council. The main purpose of this test is to screen and select overseas students to universities and colleges in English-speaking countries. However, this test is not used in USA, because in this English-speaking country TOEFL test is used to recruit and select overseas students that want to pursue their studies in this country.

Indeed, as argued by McNamara (2000, p. 7), proficiency tests "look to the feature situation of language use without necessary any reference to the previous process of teaching". Then, Henning (1987) adds that this test aids in deciding who should be allowed to participate in particular program of instruction, then it is used to screen students

with extremely low English proficiency from participation in regular university instructions. In this article, the IELTS test is reviewed, particularly focusing on the reading module. Therefore, the IELTS reading test is only the component that will be evaluated in this essay.

EVALUATION

1. THE READING TEST

In the IELTS test, as one of the assessment instrument, the reading section is tested after the listening test. The reading section emphasizes the comprehension of texts. So, in the IELTS reading test, reading comprehension is the main characteristic of texts. In this test, test-takers are required to read three reading passages consisting of 1500-2500 words. In this section, there are 38-42 questions that should be answered in 60 minutes; and the difficulty level of texts and task increases through the paper. The materials of reading comprehension are taken from magazines, journals, textbooks, and newspaper. However, all topics are in a style appropriate to candidates entering undergraduate and postgraduate course (IELTS Handbook).

Therefore, to know whether or not test-takers comprehend the text, in the IELTS reading test there are various techniques employed, they are: multiple-choice; short-answer question; sentence completion; note/summary/ diagram/flow chart/table completion; choosing from a bank of headings; identification of writer's view or attitudes (Yes/No/Not Given); classification; matching lists; and matching phrases (IELTS Handbook).

Thus, all those methods/techniques above can be categorized in Selected Response Assessments and Constructed Response Assessments, then report below evaluates and reviews those assessment instruments used in the IELTS reading test (regarding to the validity, reliability, practicality, fairness, and impact) as well as the instrument's underlying language model.

a. Assessment Instruments

Selected Response Assessments are organized based on fixed formats in which test-takers possible responses have been anticipated and then test takers are required to choose the appropriate response from those offered (McNamara, 2000). In the IELTS reading test, the methods included in Selected Response Assessments are: multiple-choice; choosing from a bank of heading; identification of writer's view or attitudes (Yes/No/Not Given); matching lists; and matching phrases.

Selected Response Assessments such as multiple-choice is regarded as the objective method in the IELTS reading test contributing to greater reliability. Alderson (2000) argues that this kind of method is built for practical reason; it is also constructed to avoid misinterpretation of some part of the text. Therefore, objective methods are used effectively to train a person's ability to think (Munby, 1968, cited in Alderson, 2000). Furthermore, other assessment is Constructed Response Assessments. Those formats require the test takers to generate a response in their own words (McNamara, 2000). In the IELTS reading test, the methods categorized into

Constructed Response Assessments are: short answer question; sentence completion; note/summary/diagram/flow chart/table completion; and classification. In those formats, test takers cannot use guessing strategy in completing the tasks. nevertheless, McNamara (2000) argues that those formats are generally more expensive to score.

b. Validity

Henning (1987, p. 89) explains that validity is the appropriateness of a given test of its component parts as an asses what it is purported to assess. Then, the text can be valid if it is supposed to test. Hughes (1989) identifies a number of different dimensions to validity; they are:

1) Construct Validity

Construct validity refers to "the extent to which we can interpret a test score as an indicator or the ability or construct we want to measure" (McNamara, 2000, p.105). In the IELTS reading test, according to construct validity, the test will be not valid if the test is not testing only reading comprehension, but also including other elements related to the reading that should not be tested such as translation or vocabulary test. Fortunately, in the IELTS reading test, those kinds of elements are not found; it totally tests the comprehension based on the text/reading material provided in that test. Therefore, the IELTS test reading test definitely follows the construct validity.

2) Content Validity

Content validity "involves inspecting the test to see if the items are valid for the testing purpose" (Richek et al, 1989, p. 132).

Content validity will be valid if it is representative for the subject or skill that will be tested. Richek et al (1989) argue that a valid reading comprehension test probably contain paragraphs with questions. In the IELTS reading test, this format can be found in this constructed response assessments such as short answer questions. Furthermore, Richek et al (1989) add that the reading comprehension test requiring test takers to match words will have questionable content validity. In the IELTS reading test, those methods are included in selected response assessments such as matching list and matching phrase. Those methods are not representative samples of the skill being tested.

In the IELTS reading test, some content are inappropriate samples from the task in the domain. When test takers are in universities, they will not be asked to read then answer in multiple-choice, sentence completion, etc, but they are asked to read for comprehension and referencing in order that they are able to do their assignment. Therefore, in content validity, McNamara (2000, p.51) argues that "judgments as to the relevance of content are often quite complex, and the validation efforts is accordingly elaborate".

Thus, in one side, regarding to the content validity, the IELTS reading test may be not valid for those reasons.

3) Criterion-Related Validity

In the IELTS reading, criterion-related validity cannot be compared to that of in the TOEFL reading test. In the TOEFL reading

test, the test is constructed based on multiple-choice questions, the test is scored more objectively, and the test is marked about 20 to 60 band score. In the IELTS reading test, the test is built not only based on multiple-choice questions (only a few questions are multiple-choice). Here, test takers cannot totally apply guessing strategy in responding to the questions, because there are also a lot of questions requiring their understanding/comprehension to answer the questions such as short answer questions. Therefore, test takers who got the band score 7 in the IELTS reading test are not equal with test takers getting the band score 50 in the TOEFL reading test. Thus, regarding to criterion-related validity, the IELTS reading test is more valid compared to the TOEFL reading test; the TOEFL reading test provides a bigger possibility for test takers to guess to answer (multiple-choice).

However, if test takers have ever taken an IELTS course, may be they often practice the reading section. In the course may be they are able to respond to questions well; probably in the real practice/test they also can respond to questions well. The score they got in the practice IELTS reading test which is similar to the score they got in the real test indicates that criterion-related validity in the IELTS reading test is good.

4) Predictive Validity

The research conducted by Dooley and Oliver (2000) shows that, regarding to predictive validity, the IELTS reading test is the best predictor of academic success as it was the only subtest of the four macro skills to

achieve a significant correlation. For example, the getting 8 of band score of the IELTS reading test will be doing the assignment well, because to do the assignment they are required to read a lot. So, reading is not a serious matter for them because they have mastered this skill which can be proved through their band score 8 in the IELTS reading test. The band score 8 will be impossible test takers got if they did not comprehend the reading well in which reading material sometimes provided the topic/text that they did not know/ not familiar before.

Therefore, in terms of predictive validity, the IELTS reading test can be valid because it shows a bigger possibility of the stability of (between) test takers' score in the IELTS reading test and their ability/score/grade when they have been already in the target university.

5) Consequential validity

The social effect of the IELTS reading test can be negative or positive. Test takers will get a negative impact if they can't obtain the target band score. Consequently, they cannot pursue their studies in the target university. (For the complete explanation will be reported in the impact part)

6) Face Validity

Related to face validity, Charge and Taylor (1996) argue that the reading module is suitable for candidates who are going to English-speaking countries to complete their secondary education, because reading module is built based on communication skills in a broad social and educational context.

On one hand, in the IELTS reading test, the content may be not representative enough for test takers that have different backgrounds. On the other hand, test takers learned the comprehension skill when they were in high school. So, probably, for some test takers, no matter what the content is, the point is that they have to comprehend the reading, but in different situation. Thus, the face validity is logical and intuitive, but usually lacks an empirical basis (Henning, 1989).

c. Reliability

Reliability refers to the stability and the constancy of test scores (Richek et al, 1989, p. 133). Hughes (1989) argues that reliability can be achieved if the test organizers or institutions constructing the test consider some aspects; they are:

1) Provide explicit and clear instructions

In the IELTS reading test, the instruction is given very clear. for example, in responding to short answer questions test takers are required to answer no more than 3 words. If this instruction is not given, probably the test-takers answer more than 3 words. This means that test takers interpret this instruction in different way, and this will not contribute to the reliability of the test.

2) Test takers should be familiar with format and testing techniques

The techniques/methods employed in the IELTS reading test such as multiple-choice, matching list, or Yes/no/Not Given format are widely used in other tests such as the test in the high school. Even test takers practice the IELTS reading test a lot before

taking the real test. so, test takers are familiar with the formats/techniques of the IELTS reading test; this contributes to the reliability of the test.

3) Using items permitting scoring which is as objective as possible

Most methods in IELTS reading test such as multiple-choice, matching heading, or Yes/No/not Given are objectives. Hughes (1989, p. 60) agrees that "since in order to respond the candidate has only to make a mark on the paper, it is possible to include more items that would otherwise be possible in given period of time". Therefore, the objective methods in the IELTS reading test provided a space to make a greater reliability.

4) Provide a detailed scoring key

In the IELTS reading test, scorer reliability is high, because the answer key is set more detail. It makes a clear acceptable answer. For example, in question number 21 (see the attachment - the example of the IELTS reading test - on the appendix pages), the answer expected is "*frames*", but if test takers answer "*frames of combs*" is also accepted. This indicates that in the IELTS reading test provides a detailed scoring key because in the answer key of the question number 21 is written "*frame (of combs)*".

Thus, the description above figures out the reliability of the IELTS reading test. However, the reliability of the test has relation to validity. Henning (1989, p. 89) mentioned that "is it possible for a test to be reliable without being valid for a specified purpose, but it is not possible for a test to be valid without first being reliable". In the IELTS reading test, it

means that reliability can be gained although such content validity is not totally valid (questionable).

d. Practicality

Practicality reflects whether the test is easy and cheap to construct, administer, score, and interpret (OLoughlin&Dyer, 2003). In IELTS reading test, it is not easy to construct the test; developers build the test in different question formats. They also organize the contents with different topics taken from different resources, and this is not easy to choose appropriate topics that will be tested. Developers must consider whether the contents are academic to be tested; this is such time consuming in constructing/developing the test. This indicates that the IELTS reading test is also not cheap to construct; it includes a lot of advanced/new materials which are needed to be revised regularly by developers.

Furthermore, related to the score, IELTS developers don't explain explicitly how the reading raw scores are converted into band levels (Deville&Turner, 2000). This made the score difficult to be interpreted. However, for developers/examiners, IELTS reading test is easy to score; most questions are based on objectives formats.

e. Fairness

Related to content and consequential validity of the IELTS reading test, probably for test takers the test is unfair, because the reading contents is not representative for their skill or background. McNamara and Rover (2006, cited in Pan, 2009) argue that test fairness should encompasses many different aspects. Bachman (1990, cited Kobayashi,

2005) adds that some facets affect the performance in the test. Related to fairness, there are some facets affecting test takers' performance in the IELTS reading test; they are:

1) Facet of the testing environment

If the testing environment such as the place for taking test is not comfortable, for example the room is very hot or not conducive for test takers, this will create an uncomfortable environment and make test takers not maximal to read. This situation may be unfair for test takers.

2) Facet of test rubric

In the IELTS reading test, test takers have to answer 40 questions in 60 minutes. It means that it takes 1,5 minutes to respond to one question. It is fair, but will be unfair because there are 3 passages of the reading that they have to read at the same time. So, 60 minutes is not enough to read/comprehend all passages as well as answer all questions.

3) Facet of the input

Visual presentation helps test takers in the IELTS reading test; they can read it several times until they understand it well.

4) Facet of the expected response

In selected response, it is fairer, because the expected answer is more objective such multiple-choice that can be scored effectively. In constructed response, maybe it is not fair enough when the answer of test takers is inappropriate (short answer is correct, but more than 3 words). However, it will be fair because in responding this question test takers cannot guess.

5) Relationship between input and response

In the IELTS reading test, test takers may use predicting or summarizing strategy to help them understand such a long passage.

f. Impact

Even though in the IELST reading test takers are not required to compose a short writing to response the questions, but still test takers can't answer the questions maximally if there are a lot vocabulary/information in the texts that they don't know; a language teacher lacks vocabulary related to engineering, a nurse lacks vocabulary related to law, so on. So, it is quite impossible for them to good band score in the reading section.

Based on the research conducted by Rasti (2009) in Iran, many test takers (56,7%) believe that reading is the most difficult section of IELTS. Therefore, many test takers can't reach the target score and they fail in the test; this is one negative effect if the IELTS reading test. For example, if test takers cannot obtain the band score 7, they will not be accepted in faculty of law in Australian universities such as Adelaide University and Deakin University.

However, the positive effect can be for the institutions. They restrict and screen the students who are not competence to be such as law postgraduate students. The faculty cannot accept the students who are not qualified which can be proved through their reading score. Then, faculty of law has good reputation in screening/selecting its overseas students. Other effect can be for the IELTS organizer such as Cambridge University or British

Council; they need to revise the reading module for a better development.

2. INSTRUMENT'S UNDERLYING LANGUAGE MODEL

In the IELTS reading test, the passage is taken from various resources. Test takers will not comprehend the text completely without having the knowledge background called 'top-down' schemata. Carrel (1983, cited in Carson&Leki, 1993) indentified 3 forms of schemata affecting test takers' performance in the reading test; they are:

a. Linguistic Schemata

Test takers are not able to comprehend to the text and its meaning without having the large vocabulary and knowing text's structure. Therefore, language knowledge is urgently needed by test takers.

b. Formal Schemata

Test takers can use prediction strategy to respond to 'sentence completion' or 'clarification' question in the IELTS reading test.

c. Content Schemata

If test takers had the knowledge of the reading content being tested, they can comprehend text well. They can use skimming or scanning strategy to find the information needed; they have not to read several times because they have known the topic/information called topical knowledge. Bachman and Palmer (1996, cited in Fulcher&Davidson, 2007) add that the affective or emotional included in personal characteristics correlate of topical knowledge determines whether test takers will engage with a particular task in the test.

Therefore, characteristics of the language use or test task environment cover the topical and language knowledge including its affect, then those all components are controlled by the strategic competence. By analyzing all schemata and the strategy involved in the IELTS reading test, the components of language use and language test performance mentioned by Bachman and Palmer (1996) can be identified as shown by the diagram below.

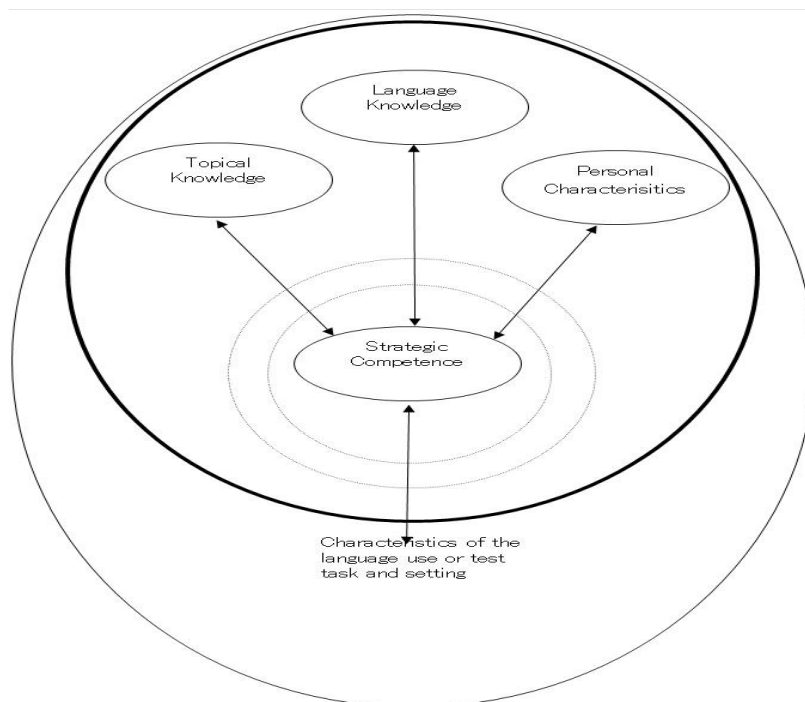


Figure 1: Some components of language use and language test performance (Bachman and Palmer 1996: 63), cited in Huuskunen and Kahkonen, 2006, p. 9.

Thus, instrument's underlying language model of the IELTS reading test is based on the 1996 model provided by Bachman and Palmer.

CONCLUSION

The purpose of IELTS is to select and screen international students that want to pursue their studies in British and Australian universities. Regarding to validity, some dimensions of validity such as content validity, consequential validity, and face validity are less valid. In term of reliability, the IELTS reading test provides a greater reliability of the test.

However, the IELTS reading test is not totally practical and fair. The IELTS reading test is not practical because some methods included in constructed response assessments such as short-answer question and sentence completion are generally expensive

to score. Then, the test will be unfair when some facets have negative effects to test takers; so, the IELTS developer should encompass many aspects such as facets affecting test takers' performance in the IELTS reading test in order to make the IELTS reading test can be valid in all dimensions.

Moreover, in the IELTS reading test, for a better development, developers need to considers test takers' educational background when revising the content of the reading passage; this can minimize the negative impact such as test takers fail to obtain the target band score.

Finally, many researchers investigated the impact of the test to test takers as individuals; therefore further researchers are urgently needed to examine the IELTS reading test impact or wash back to educational system.

REFERENCES

- Alderson, J. C. (2000). *Techniques for testing reading: Assessing reading*. Cambridge, UK: Cambridge University Press, pp. 202-270.
- Carson, J., & Leki, I. (1993). *Reading in the composition classroom: Second language perspective*. Boston: Heinle & Heinle.
- Charge, N., & Taylor, L. B. (1996). *Recent developments in IELTS*. Retrieved 23 August, 2010, from <http://eltj.oxfordjournals.org/content/51/4/374.full.pdf>.
- Deville, M. C., & Turner, C. E. (2000). *What to look for in ESL admission test: Cambridge certificate exams, IELTS, and TOEFL*. System journal, vol. 28, pp. 323-329. PERGAMON Database.
- Dooley, P., & Oliver, R. (2002). *An investigation into the predictive validity of the IELTS test as an indicator of future academic successes: IELTS Annual Review*. Prospect Journal, vol. 17, no. 1, pp. 36-54.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Abingdon, UK: Routledge.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Boston: Heinle & Heinle.
- IELTS Handbook (2007). *Prepare for IELTS skills and strategies: Book two; Reading and Writing*. Sydney: INSEARCH University of Technology Sydney.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.
- Huuskonen, M. L., & Mirva Kahkonen, M. (2006). *Practicing, Testing, and Assessing Oral Skills in Finnish Upper Secondary Schools: Teachers' Opinions*. Department of Languages. University of Jyväskylä, Retrieved 23 August 2010, from https://jyx.jyu.fi/dspace/bitstream/handle/123456789/7375/URN_NBN_fi_jyu-2006384.pdf;jsessionid
- Kobayashi, W. (2005). *An investigation of method effects on reading comprehension test performance*. Lifelong learning: Proceeding of the 4th Annual JALT Pan-SIG Conference. Tokyo: Keizai University.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- OLoughlin, K., & Dyer, W. (2003). *Best Practice in Classroom Assessment: From rhetoric to reality*. 16th Educational Conference Melbourne 2003. Retrieved 20 August, 2010, from http://www.englishaustralia.com.au/ea_conference03/proceedings/pdf/030F_OLoughlin_Dyer.pdf
- Pan, Y. C. (2009). *The impact of test design on teaching*. The International Journal of Education and Psychological Assessment, vol. 3, pp. 94-104. Time Taylor Academic Journals.
- Rasti, I. (2009). *The Iranian candidates' attitudes towards IELTS*. The reading module. Asian EFL journal, vol. 11, no. 3, pp. 110-155.
- Richek, M. A., List, K. L., & Lerner, J. W. (1989). *Reading problems: Assessment teaching strategies*. New Jersey: Prentice-Hall.

Shaw,S,&Falvey,P. (2008). *The IELTS writing Assessment Revision Project: Towards a revised rating scale*. University of Cambridge, ESOL Examinations. Retrieved 12 August 2010, from http://www.cambridgeesol.org/assets/pdf/research_reports-01.pdf