

A Probe into the different aspects of ‘Validity’ and ‘Reliability’ of IELTS writing test

Hosne Al Noor

Lecturer, Center for Language Studies, University of Liberal Arts Bangladesh (ULAB), Dhaka, Bangladesh

Abstract— The current paper reviewed the writing section of IELTS. Writing module of IELTS is a test with good face validity. Issues like developing the test content, elicitation of writing samples from the test takers, assessing those writing properly and many other factors are involved in the entire process of IELTS writing test (Mickan, 2003). While developing the test content, a test developer needs to ensure that the content is suitable for the test takers. This test also demands well-trained markers and focused testing of the writing skill. The study reviewed the different significant aspects of the validity and reliability of IELTS writing test. Finally, some recommendations and suggestions were provided to increase the validity and reliability in a more credible way.

Keywords— IELTS writing test, validity, reliability, content development, test administration, test instruction.

I. INTRODUCTION

IELTS is a prevalent language proficiency test of English all over the world. This test is used to evaluate the proficiency level of the learners in their use of English language. IELTS is consisted of 4 modules and these modules are: reading, writing, listening and speaking. Each of the skills is imperative and requires different types of attention and concentration to get a better result. However, the current study deals with the writing skill. There are different aspects of IELTS writing test that makes this module an interesting one to review.

Score of IELTS writing test can have far reaching effect in the life of a test taker as every test taker appears for this test with a purpose. After passing the 12th standard, the author wanted to go USA for undergraduate degree. So, the author appeared in the IELTS test and received a band score of 7.00 with a score of 7.5 in the writing module. Then the author completed bachelor degree in English and appeared for IELTS test again after a gap of 4 years. This time, the band score was up to the expectation and but the researcher received a score of 6.5 in the writing section that made him surprised. Despite of doing bachelor degree in English and getting favorable exposure to improve his English, the author scored poorly in the IELTS exam. However, the author felt that his English was much more improved after the bachelor degree than it was after the 12th standard. This incident

inspired the author to review the writing module of IELTS test.

II. OBJECTIVES

In this review paper, the author investigates the areas in which this particular test has rooms for improvement. The researcher also focuses on the good sides of this test. The major aim was to review the validity and reliability of IELTS writing test.

III. METHODOLOGY

Bachman and Palmer (1996) provided a test usefulness framework in which they talked about validity, reliability, authenticity, interactive-ness, washback and practicality. A test can be reviewed on the basis of these criteria. However, the researcher here has deployed the validity and reliability aspects of test usefulness to review the IELTS writing module. The researcher deployed a natural descriptive approach with conceptual and relational analyses to review the IELTS writing test from the perspectives of the different aspects of validity and reliability.

IV. LITERATURE REVIEW: GENERAL INFORMATION ON IELTS WRITING TEST

Before reviewing a test, it is important to ensure that the readers have clear idea about the test. IELTS writing tests are of two types. They are called: Academic and General Training (GT). Academic IELTS writing test is taken for academic purpose and GT IELTS writing test is taken for job and immigration purpose. IELTS writing test has a duration of 60 minutes and it is consisting of two type of tasks. In task 1, the test takers perform a diagram analysis in about 150 words. GT candidates write a letter in place of diagram analysis for their task 1. However, Task 2 is same for both GT and academic module. Here, the test takers write a composition in about 250 words. Neither of the tasks (1 and 2) provide any option for the test takers.

Both academic and GT module writing test are evaluated on a scale of 9. The score received by a test taker symbolizes his level of proficiency in writing English. Task 2 of the test carries more weight than task 1 and the scripts are assessed by the trained markers (Uysal, 2010). Markers use rubric in assessing the scripts.

V. FINDING AND DISCUSSION

5.1 Different aspects of validity of IELTS writing test

5.1.1 More validity in content development

In general sense we can say that IELTS writing test has content validity as the items provided in the test make the test takers write and this test also ensures that the writing sample of the writers are also large enough to evaluate their writing. However, test developers still have some scope of improving the content validity of the test.

In the academic module of IELTS writing test, the test takers are given the task of analyzing a diagram. There are different types of diagrams that are given in the test, such as, pie chart, bar chart, picture description etc. This variation in diagram hampers the content validity of the test to some extent. There can be certain candidates who has good writing skill but they might find it difficult to understand pie chart. Pie chart or bar chart demands a certain level of analytical ability from the candidates. Moreover, in describing pictures there can be series of pictures that represents a particular process. This type of task can cause ambiguity and it might affect candidate's writing performance.

In both academic and GT module, the candidates need to write an essay (mostly argumentative) in their task 2.

This task provides detail instructions and sometimes provides a context for the essay as well. However, the topic of the essay is very crucial as the test developers need to make sure that every test taker has fair idea about the topic of the task. There are certain topics that demands prior knowledge of the test takers. In such scenario, the focus of the test can be shifted from testing the writing skill to testing the knowledge of the candidates. For example, air pollution can be an essay topic and the test developer may feel that it's a general issue that everyone has some understanding. However, we can't expect that every candidate has required knowledge to write an essay on air pollution.

One major characteristics of a valid test is that it restricts the candidates to test the target language ability (Hughes, 2007). However, too much restriction on the test takers can also minimize the content validity of the test. Test takers don't get any option to choose from in the task 1 and task 2 and it puts them in a challenging situation where their level of knowledge also get tested along with their writing skill.

Another important point is, test developers only uses essay and diagram analysis in academic IELTS writing test. Lack of variation in the test items help the candidates to get strategic and thus it gets more difficult to test the writing ability of the test takers. Moreover, difficulty level of task 1 and task 2 don't correspond to each other. It is seen that task 2 is more high demanding than task 1 and many candidates performs better in task 1 but fail to do so in task 2 (Nguyen, 2015). According to Hughes (2007), it is important to ensure content validity of a test in order to ensure positive washback effect on both raters and test takers.

5.1.2 Validity in scoring

If a test can measure the abilities it claims to measure and the test takers find the test relevant and useful to test the intended target ability then we can say that the test has face validity (Brown & Abeywickrama, 2004). It can be said that IELTS writing test has good face validity as it is a writing test and the writing ability of the candidates get tested. For example, the candidates are supposed to write an essay in this test. In place of essay writing, if they were given a grammar task, still they would write something but that particular test would be more useful to judge the grammatical knowledge than the writing skill. It would have minimized the face validity of the writing test.

A test can have valid content but if the scoring procedure is not valid then it can minimize the reliability of

the test. While assessing the writing ability, if the rater put too much emphasis on spelling, punctuation and grammar then it will hamper the scoring validity of a writing test. These mechanical features need to bring into consideration but primary focus has to be on the writing skill of the candidate. IELTS writing test ensures good validity in scoring by giving importance on Task Response (TR) and Coherence and Cohesion (CC) of the candidate (Soleymanzadeh&Gholami, 2014). However, they should expose the detail scoring of writing to each candidate rather than giving them just a numerical score.

5.1.3 Criterion related validity of IELTS writing test

According to Hughes (2007), criterion related validity refers to the extent to which the test is able to assess the target ability of the candidates. Hughes (2007), refers to two types of criterion related ability, which are: concurrent validity and predictive validity. IELTS writing test has strong concurrent validity but the predictive validity of this test is still questionable.

IELTS writing test shows good evidence of concurrent validity as this test cover almost all the aspects of a candidate's writing ability that need to be tested. Usually the aspects that a writing test covers are: a candidate's sentence construction style, organization ability, grammatical accuracy, spelling and punctuations. IELTS writing module testes all these features in a 60 minutes test by ensuring the collection of large enough sample of writing from the test takers. There are writing tests that can take up to 3-4 hours but the IELTS test is designed in such a way that it can provide an equally valid evaluation with a test that has a duration of 60 minutes. IELTS writing test is able to provide an estimation of the writing ability of the candidates and thus it provides good concurrent validity (Uysal, 2010).

Lack of strong predictive validity is a significant setback for IELTS writing test. Moore and Morton (2005), compared IELTS writing task 2 with 155 academic essays written by Australian university students. The result of the study showed that IELTS essay writing belongs to the non-academic genre and task 2 is not appropriate to judge the writing ability of the students. The author's practical experience also matches with claim of Moore and Morton. People do exceedingly well at the postgraduate level even after scoring 6.00 in IELTS academic writing test. On the other hand, students who are holding a score of 7.5 in IELTS writing seems to struggle in getting good mark in their academic essays due to their writing style. IELTS task 2

demands candidate's knowledge, ideas and experience about the essay topic but academic essays demand subject related specific knowledge (Morton, 2007). Task type of IELTS test reduces its predictive validity to some extent.

In order to get the desired outcome from a writing test, it is important to ensure the validity of the test. Hughes (2007), talked about different issues that need to be considered while developing a valid test. Some important points that are needed to consider while validating a test are:

- a. Test content have to be selected from representative sample.
- b. Test instruction need to be clear and detailed.
- c. All the items of a writing test should correspond in their difficulty level.
- d. Nature of scoring should reflect what is being tested.
- e. Direct testing of the candidates with the assurances of collection of long enough samples.

5.2 Different aspects of reliability of IELTS writing test

Reliability is a very important test quality. A test can ensure validity only when it is reliable. Reliability is a feature that influences other test qualities as well. A reliable test can have positive washback on both test taker and test developer by ensuring authentic and interactive test content and testing procedure. Reliability refers to the consistency in measurement (Brown &Abeywickrama, 2004). If a particular test can bring consistent and dependable outcome irrespective of the group of test takers or the test setting that is when we can consider that particular test as a reliable one. There are certain aspects of reliability which are ensured by IELTS writing test but there are a few aspects of this test that demands more reliability.

5.2.1 More consistency in rater reliability

According to Rezaei and Lovorn (2010), rubric based writing evaluation can bring the desired outcome but we also need to ensure that the teachers are well trained in the use of rubric. Raters of IELTS writing test use rubric in order to unsure rater's reliability. Rubric helps them in categorizing different aspects of writing and thus they can score the scripts in a consistent manner. Raters are supposed to consider language elements like writing style, accuracy, grammar, punctuation separately and then mark those different elements on the basis of the importance of every single element. However, scoring a script is a complex decision-making activity and often the raters mark a script by

considering the full text at a time (Mickan, 2003). It is seen that sometimes the markers use their individual perspective while scoring an IELTS writing script. Sakyi (2000) talked about different type of reading behavior of the readers and emphasized the fact that different raters can focus on different aspects of writing while scoring a script. Few might focus on the writing error of the writer, whereas few might focus on the informativeness of the text. Rater's personal reaction to the topic of the text can also determine the score of the writer. Factors like these make the inter-rater reliability of the test questionable. Use of rubric and teachers training can ensure intra-rater reliability to some extent but inter-rater reliability is still not at the acceptable level as the markers don't provide any written feedback of the scripts either. According to Weigle (2007), teacher training plays a beneficial role in developing evaluation skills of the markers but individual perception of a marker is based on his personal belief and that perception can influence the scoring process.

Electronic scoring system can be a possible solution to ensure inter-rater reliability. Every electronic scoring system contains a large sample of writing. If the scripts are checked electronically, it will not only restrict the implementation of individual perspective of the raters but also will save a lot of time (Dikli, 2006).

5.2.2 Reliability in Test administration

The condition in which a test is administered can hamper the reliability of a test (Brown and Abeywickrama, 2004). We can say that IELTS writing test ensures administration reliability. IELTS writing test is conducted along with the reading and listening test on the same date. Usually the test authority chooses test venues where a good number of students take part in the test at a time. Students receives papers that contains instructions of the task and the students are supposed to write their answer on the space provided on the paper. Students are given the scripts for writing after the completion of the listening test. In between both the test, test takers get around 5 minutes to settle themselves down. The exam invigilators ensure constant supervision and the test takers can use either pen or pencil to write down their answers. IELTS authority ensures that there's no loud noise around the exam center. This test also maintains reliability in test administration by providing clear photocopies of questions and comfortable sitting arrangements for the students.

5.2.3 Student related reliability of IELTS writing test

Test pattern, time challenge of the test, exam setting can have positive or negative effect on the test takers. From personal experience of being a test taker, the researcher has seen that the test pattern and the test setting can cause anxiety for the candidates. These factors often work as the reason behind the difference between the actual score and true score of the candidates (Hughes, 2007). Sometimes it is seen that one particular examinee takes the IELTS writing test twice within a period of 1 month and the score received by him varies significantly. IELTS test consider test-retest reliability and they try to provide questions with the same difficulty level for each test. So, it is evident that the student related reliability plays a crucial part in determining the score of the candidates.

5.2.4 Reliability in Test instruction

Candidate's performance in a writing test depends largely on the understanding of their test instruction. Due to lack of understanding, even a good writer can write something which is not relevant to what was asked in the question and it can result in receiving poor score. Task 2 of the IELTS writing test provides a context for the essay where a test taker can read the instructions and then can choose a side (of the topic) to defend in his argumentative essay. However, the instruction of the task 1 of IELTS academic writing test can prove to be ambiguous for the candidates. People with average analytical ability might find it difficult to understand the instruction clearly (O'Loughlin and Wigglesworth, 2003). Moreover, sometimes students are given the task of describing a series of picture (process diagram) in their task 1, such task uses arrows and other signals that might confuse the test taker and there's always a possibility that he might misinterpret the pictures. So, the lack of clear and explicit instruction in the task 1 of the test is a factor that hampers the reliability of the test.

VI. CONCLUSION AND RECOMMENDATION

IELTS writing test is considered as a standardized writing test that evaluates the proficiency level of the candidates (Soleymanzadeh&Gholami, 2014). This test has been acceptable level of validity and reliability in test content, test instruction, scripts evaluation. However, there is still room for improving the validity and reliability in many aspects of this test. This review paper is based on the previous works of the other researchers and the personal experience of the author. The researcher feels that a

qualitative research on this topic can help even more to get the better idea about where we need to increase the validity and reliability of IELTS writing test. From personal evaluation, the researcher feels the following measures can be taken to improve the validity and reliability of the test:

- a. More variation can be brought in selection of the test content. We can add different items to minimize the predictability of test item.
- b. A sample answer can be provided in the scripts in order to ensure the better understanding of the test by the candidates.
- c. Every script can be checked twice to increase rater reliability. We can use a combination of scoring by using rater and electronic scoring procedure. It will ensure that the perception of rater won't have much effect on the score of the candidates. However, use of technology has to deal with the practicality aspect of the test.
- d. Along with the numerical score, written feedback can also be provided. It will help the test takers to understand their strengths and weaknesses.

REFERENCES

- [1] Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Vol. 1, Oxford University Press.
- [2] Brown, H. D., & Abeywickrama, P. (2004). *Language assessment. Principles and classroom practices*. Longman.
- [3] Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment*, 5(1), Retrieved [January 30, 2019] from <http://www.jtla.org>
- [4] Hughes, A. (2007). *Testing for language teachers*. Cambridge: Cambridge University Press.
- [5] Mickan, P. (2003). 'What's your score?': An investigation into language descriptors for rating written performance. *International English Language Testing System (IELTS) Research Reports 2003*, 5, 125.
- [6] Moore, T., & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes*, 4(1), 43-66.
- [7] Morton, J. (2007). Authenticity in the IELTS Academic Module Writing Test: A comparative study of Task 2 items and university assignments.
- [8] Lan, N. T. (2015). The effect of task type on accuracy and complexity in IELTS academic writing. *VNU Journal of Science: Foreign Studies*, 31(1).
- [9] O'Loughlin, K., & Wigglesworth, G. (2003). Task design in IELTS academic writing task 1: The effect of quantity and manner of presentation of information on candidate writing. *International English Language Testing System (IELTS) Research Reports 2003: Volume 4*, 89.
- [10] Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15(1), 18-39.
- [11] Sakyi, A. A. (2000). Validation of holistic scoring for ESL writing assessment: How raters evaluate compositions. *Fairness and validation in language assessment*, 129-152.
- [12] Soleymanzadeh, L., & Gholami, J. (2014). Scoring argumentative essays based on thematic progression patterns and IELTS analytic scoring criteria. *Procedia-Social and Behavioral Sciences*, 98, 1811-1819.
- [13] Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT journal*, 64(3), 314-320.
- [14] Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, 16(3), 194-209.