

Validity and Fairness of TOEFL iBT Reading Test

Miaomiao Sun

Baoji University of Arts and Sciences, Baoji 721000, China.

Abstract: The validity of the TOEFL, a large-scale worldwide test and a high-stake test, is the most concern for test-takers. The validity of TOEFL scores has been found positive as a whole but in reading comprehension. Four aspects are analyzed to study the validity of fairness of TOEFL iBT reading test. Results show that reading score are under-represented, with some cognitive validity issue, negative washback, and unfairness.

Keywords: TOEFL iBT Reading Test; Validity; Fairness

1. Introduction

As a language proficiency test, the Test of English as a Foreign Language (TOEFL) has established itself as a large-scale worldwide test over the years. TOEFL test is also a high-stake test since the score of TOEFL will be strong evidence of the test takers language ability. Given these characteristics, the validity of the TOEFL test is the most concern. Studies have constantly been conducted on the validity of TOEFL scores to interpret test takers English language proficiency, and consistent and positive results have been shown on the test as a whole and in other domains but in reading comprehension. The reading section assesses test takers academic reading skills using multiple-choice items. Evaluated through the reading test are three academic reading skills: basic information skills, inferencing skills, and reading-to-learn skills. A wide variety of academic subjects are covered in the testing passages, ranging from astronomy to anthropology. Normally, the reading section involves three passages; occasionally, it includes an extra passage, or an experimental section of reading. In sum, TOEFL iBT reading test is characterized by the following: (1) internet-based testing format; (2) a test of academic reading skills; (3) broad academic subjects; (4) an experimental section^[1].

2. iBT testing mode: concurrent validity

First introduced in 2005, TOEFL iBT has experienced a constant comparison with the previous paper-based test (PBT) and computer-based test (CBT). Some advantages of TOEFL iBT are obvious: more effective scoring process and detailed score report. With the widespread internet, students become accustomed to reading articles and books online since the internet provides relatively convenient access to substantial reading materials. Also, while reading, they often take some important notes, which is allowed during the test. Thus, considering the online reading and permission for note-taking, the iBT mode is quite parallel to the real-life situation. But using iBT mode to test reading comprehension might face some concurrent validity issues. Research have shown that a large difference in the reading exists between PPT and CBT on a large-scale statewide examination (Kim & Huynh, 2008). Even though there is no significant difference in other domains under CBT, a worse performance is noticed in reading comprehension. Reading scores were found lower for scrolling in CBT than for paging in PPT (Pommerich, 2004). Given that iBT mode is conducted similarly to CBT mode, a similar situation could happen on the TOEFL iBT reading test. When two different testing administration modes do not correspond to each other, there will be a concurrent validity issue to that test. In light of the above, the TOEFL iBT reading test might have some valid issue in the concurrent aspect^[2].

3. Test of academic reading skills: construct validity and cognitive validity

Evaluated from the reading construct, TOEFL iBT reading test is quite valid since it involves elements such as identifying main and supporting ideas, relating cause to effect, recognizing the rhetorical functions performed by sentences in the development of a text, and understanding different types of the meaning of words/semantic associations. Vocabulary questions aim to elicit a test takers ability to infer the meanings of individual words and phrases used in the passage. Lims (2014) study showed that one third of the participants chose the vocabulary questions correctly without reading the passage. More importantly, the washback of vocabulary items might not be positive^[3]: obviously, some test takers would pre-memorize as many academic words as possible rather than practice as much their inference skills from the context. In addition, R2L questions mainly assess a test takers metacognitive reading skills. Using the Think-Aloud approach to analyze the participants response to R2L questions, Cohen and Upton (2006) reported that subjects were summarizing in their heads what the key ideas were and then moving to the test item to find the answer that matched the understanding they had in their heads; in this way, the R2L items are viewed as testing tasks rather than reading tasks since little mental framework have been involved. iBT reading is like a test-taking task with academic-like aspects to it. Thus, evaluated from the aspect of testing academic reading skills, TOEFL iBT reading test faces cognitive issues in some of its question items^[4].

4. Broad academic subjects: content validity

Covered in the TOEFL iBT reading test are highly variable academic subjects, including astronomy, geology, biology, anthropology, humanity, history, and arts, each of which is expected to be read and learned in the university. It is an ideal situation that a test involves as large a content of materials as it is supposed to. In reality, however, it is impossible to do so. ETS may adopt sample widely and unpredictably policy, meaning to cover a broad range of content and, most importantly, in an unpredictable way. Thus, it is not easy for test-takers to prepare the content before the actual test. Seen from the content perspective, TOEFL iBT reading test is quite valid. However, this sampling widely and unpredictably may result in inconsistent reading scores for those who rely on content knowledge to understand and answer the item questions^[5]. As background knowledge cognitively affects reading comprehension. It is easy to predict that if a test taker fails to process the text, he or she will use his or her background knowledge to understand the text. If this happens, the test itself cannot be considered cognitively valid since it is the test takers schema rather than reading ability that decides the testing results.

Many studies have been conducted to look for the correlation between background knowledge and reading comprehension. For instance, Awabdy (2012) found that that a significant association was shown between background knowledge and reading comprehension in those passages that discussed specialized knowledge of a topic. Clapham (1996) noticed that the medium-level students heavily use their background knowledge to understand the text. Negative washback can also be predicted: medium-level students desiring higher scores might spend more time learning various academic subjects than practicing their English reading skills^[6]. Given this situation, the content knowledge will hurt reliability of the score. Also, testing the ability to read various academic subjects is unfair to test takers since native English speakers who apply to the same university are not required to demonstrate the same reading ability.

5. The experimental section: fairness

An extra reading passage is a distinctive characteristic of the TOEFL iBT reading test. This happens constantly on every TOEFL iBT reading test but unpredictably to individual test-takers. In other words, only when the reading section begins will a test taker know whether he or she will be tested additionally. The extra reading passage is randomly placed, and the extra items look the same as graded questions, so test-takers cannot guess which passage is the extra one. The extra questions are either questions used to make test scores comparable across administrations or new questions to help determine how such questions function under actual testing conditions. It sounds reasonable; however, the truth is that the extra reading passage is experimental research to determine whether the new reading material is reliable and valid. So, it seems that test takers are taken advantage of as pilot subjects in a high-stake examination. Even though the extra items are not scored and will not influence the reading score, the experimental reading section makes the reading test very long^[7]. Spending 72 minutes rather than 54 minutes to focus on reading may not be easy for some test takers throughout the reading test. Also, if the

experimental reading passage is more difficult than the other three reading passages, test takers may become anxious and thus influence their reading scores. Therefore, compared with other test-takers, these pilot subjects are unfairly treated in the TOEFL iBT reading test.

6. Conclusion

As an assessment can only be considered valid for a certain purpose (Green, 2013). The TOEFL iBT reading test is valid in some aspects except in some others. Noticed are the under-representation of the reading score due to the testing mode, cognitive validity issue of some question items, content validity issue of broad academic subjects and its negative washback to medium-level students, and unfairness of the experimental section to the piloted test takers.

References:

- [1] Awabdy GW. Background knowledge and its effect on standardized reading comprehension test performance[M]. University of California, Berkeley, 2012.
- [2] Clapham C. The development of IELTS: A study of the effect of background on reading comprehension[M]. Cambridge University Press, 1996.
- [3] Cohen AD, Upton T A. Strategies in responding to the new TOEFL reading tasks[J]. ETS Research Report Series, 2006, 2006(1): i-162.
- [4] Green A. Exploring language assessment and testing: Language in action[M]. Routledge, 2013.
- [5] Kim DH, Huynh H. Computer-based and paper-and-pencil administration mode effects on a statewide end-of-course English test[J]. Educational and Psychological Measurement, 2008, 68(4): 554-570.
- [6] Lim HJ. Exploring the validity evidence of the TOEFL IBT reading test from a cognitive perspective[M]. Michigan State University. Second Language Studies, 2014.
- [7] Pommerich M. Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests[J]. The Journal of Technology, Learning and Assessment, 2004, 2(6).

About the Author:

Miaomiao Sun (1986.12-), Female, Han nationality, Shaanxi Baoji people, teaching assistant, Master, Research Interests: Applied Linguistics and English Education.