Title of Project:
Exploring the Validity Evidence of the TOEFL IBT
Reading Test from a Cognitive Perspective

Researcher:
Hyo Jung Lim
Michigan State University
hyojung@msu.edu

Research Supervisor:
Dr. Paula Winke

Project Summary:
The current study explores the validity of the TOEFL iBT reading test by examining test takers’ reading processes. The study consists of three sub-studies led by the following research questions:

1) To what extent does the TOEFL iBT reading test measure the reading construct?
2) To what extent does the TOEFL iBT reading test activate the cognitive processes being measured?
3) To what extent does test format affect students’ test scores and test-taking processes?

The eye-tracking study, which was conducted to answer the second and third research questions, aimed to reveal whether the high-stakes, standardized second/foreign language reading test elicits the same type and level of cognitive processing from test takers as real-world academic reading tasks and whether test takers actually deploy required reading skills as assumed. I believe this investigation could be a meaningful, empirical attempt in that the study takes test takers’ minds under consideration for the purpose of test validation. To date, test validation in language testing has heavily relied on correlation-based statistical analysis, which only concerns test outcomes. This statistical approach is limited because numbers do not carry conceptual information, and test scores do not inform us about how test takers derive their answers. That being said, the cognitive processing approach can complement the conventional, statistic-based test validation methods while illuminating the test takers’ processes during a test.

Ninety Chinese ESL students attending a large Midwestern university participated in the experiment. The participants took two reading tests, one with multiple-choice questions and the other with open-ended questions, on a computer whose screen was fitted with a Tobii TX300
eye-tracker (www.tobii.com). Their eye movements were recorded while taking the test. Participants also completed a vocabulary test, a grammar test, a lexical processing task, a sentence processing task, a working memory task, a strategy questionnaire, and a stimulated recall interview. The collected data were analyzed in three different ways to answer the above research questions.

The first analysis aimed to investigate the knowledge source or skills of learners that the TOEFL iBT reading test taps into. The reading comprehension scores were regressed on vocabulary knowledge, grammar knowledge, word processing skill, sentential processing skill, working memory, reading strategies, test management strategies, and test-wise strategies. The results showed that vocabulary and grammar knowledge and word processing skills were the most important predictors. Learners’ use of test-wise strategies was, however, negatively correlated with test scores; that is, those who used test-taking tricks were likely to score low.

The second analysis aimed to examine learners’ reading processes by tracking their eye movements under testing conditions. First, the predetermined criteria of Bax and Weir (2012) were used to analyze high scorers’ eye recordings (e.g., gaze plots). Then, successful readers’ eye movements (e.g., total fixation duration) were compared to those of unsuccessful readers item by item, as in Bax (2013). To identify reading types, reading time and angles between fixation points were considered. The following patterns emerged: expeditious reading skills (i.e., those skills used to read text quickly) were rarely activated, especially when skimming across paragraphs was anticipated (e.g., learning to read questions). Instead, careful reading (e.g., reading to learn or reading to memorize) at a sentential level seemed to prevail. For vocabulary questions that aimed to measure learners’ abilities to make inferences, test takers seemed to rely on their stored knowledge. In regard to the comparisons of successful and unsuccessful readers, successful readers were not always faster at reading and locating key information. In cases in which group differences appeared in the eye-movement comparisons, the item types were often limited to vocabulary and factual questions, which primarily demand locating and matching skills at the lexical level.

The third analysis was aimed at inspecting the test format effect on test-taking strategies and test scores. Given that learners’ use of test-taking strategies is induced by test format (Rogers & Harley, 1999), the presence of a test format effect was assumed to evidence the impact of test-taking strategies on test results. First, the reading test scores were regressed on the test format, the text effect, and the group difference factors. The results showed that students were likely to score higher on the multiple-choice questions than on the open-ended questions. Then, the eye movements of the participants who received a full score on the open-ended questions were compared to those of the participants who correctly answered the corresponding multiple-choice questions. The eye-tracking analysis showed that it took more time for test takers to read a text, a
question stem, and the key phrases in the text when open-ended questions were given. Additionally, the presence (or absence) of response choices seemed to alter the focus of test takers’ attention on the text. Strategic reading behaviors were more evident for the multiple-choice questions than the open-ended questions.

The results of the three different analyses did not consistently support the validity of the TOEFL iBT reading test. Given that learners who had a large vocabulary, advanced knowledge of grammar, and rapid recognition likely scored higher than those who did not, the test seems to properly tap into learners’ construct-relevant cognitive resources. However, careful scrutiny of test takers’ reading processes revealed that, in general, the test items failed to elicit learners’ expeditious reading skills, which contradicts the original intention of the test. Second language reading researchers have voiced their concerns about the importance of L2 reading fluency in the post-secondary academic setting (e.g., Grabe, 2010). That being said, the current study provides some issues for language testers to ponder upon, including how to realize the assessment of learners’ fast reading skills in high-stakes, standardized tests especially at the global level and the extent to which test tasks can (or should) mimic real-world academic reading tasks. With regard to the effect of test takers’ strategies on test performance, the first analysis suggested that there was little or no effect. However, the test format effects found in the third analysis alluded to the influence of learners’ test-taking strategies, specific to the multiple-choice questions, on their test-taking processes and test scores.

The study neither confirms nor discredits the validity of the TOEFL iBT reading test. Rather, it emphasizes the need to more closely examine learners’ test-taking processes for test validation rather than only examining the outcomes. As the eye movement data show, test takers do not necessarily take the intended path to capitalize on the intended cognitive resources. Bachman (1990) argued that all types of validity complement one another, each providing different pieces of evidence to justify test scores or uses. Thus, various types of empirical (both qualitative and quantitative) methods should be encouraged for test validation.
References


Overmann, M. (1999), Reading is guided creation. *Fremdsprachenunterricht, 43*(5) 327-332.


