

---

# A Review of the Reading Section of the TOEIC

*Manami Suzuki and Carolina Daza*

---

## *Introduction*

In 1979, the Educational Testing Service (ETS) developed the TOEIC (Test of English for International Communication), an English proficiency test for people working in international environments, based on a request from the Japanese Ministry of International Trade and Industry. The Chauncey Group International, a subsidiary of ETS, currently develops and publishes the test. Over two million people per year take the TOEIC ([www.toeic.com](http://www.toeic.com)). According to the *TOEIC Report on Test-Takers Worldwide, 1997-98*, 63% of the TOEIC results were used in Japan, 29% in Korea, and 8% in other countries.

Most reviews of the TOEIC have been descriptions of the test (Gilfert, 1996; Perkins, 1987). The TOEIC comprises the listening and the reading section. Buck (2001) reviews only the listening section. For the reading section of the TOEIC we could find only one critical review (Richards, 1992) published over the two decades since the test was developed. Therefore, our purpose in this article is to review critically the reading section based on recent studies of language assessment, particularly for construct validity and content validity, which are considered by language testing researchers (Bachman, 1990; Cumming, 1996) as fundamental for validation of language tests.

## **Potential Users and Purpose**

The *TOEIC User's Guide* suggests that the test can be used by corporations, English training programs, English language schools, or individuals. The guide reports that more than 4,000 corporations around the world have used the test for recruiting, promoting, and deploying employees. The test is also used by corporations for screening employees for several purposes: (a) technical training in English; (b) overseas assignments; and (c) language training. In addition, corporations can use the test for diagnosing their language training programs and employees' English proficiency. Educational institutions in both ESL and EFL contexts have adopted the TOEIC for formative or placement purposes (Gilfert, 1996). The TOEIC has been adopted by universities in Japan and Korea for the formative assessment of students' English proficiency and to assist students with future employment (Gilfert, 1996; Kim, 2001). Finally, the TOEIC has been taken by individuals who wished to show a measure of their English proficiency to their potential employers.

## Format of the Reading Section

The reading and listening sections can be administered and scored independently. Examinees are to complete the reading section in 75 minutes. The section comprises the last three parts of the test, Part 5 (Incomplete Sentences), Part 6 (Error Recognition), and Part 7 (Reading Comprehension). Parts 5 and 7 have 40 questions each, and Part 6 has 20 questions. Altogether, the section has 100 multiple-choice questions with four possibilities each. The scores are measured by the number of correct responses, converted to a number on a scale from 5 to 495 in intervals of 5 points.<sup>1</sup> The scores of the listening section are also converted to a number on a scale of 5 to 495. Adding the listening and the reading scores together is the TOEIC total score on a scale ranging from 10 to 990.

## Issues of Construct Validity and Content Validity

In this section we review construct validity and content validity of the reading section and consider the validity of the entire TOEIC test.

### *Construct Validity*

The TOEIC claims to assess English for International Communication, but has no clear identification of the TOEIC test domain. For example, the *TOEIC Examinee Handbook* (1998) says, "TOEIC does not test 'business English'" (p. 12), whereas the handbook explains that the TOEIC is a test that measures "English with others in business, commerce and industry" (p. 1). The handbook also states that the TOEIC assesses "general English proficiency in an international environment" (p. 12) and "the everyday English skills of people working in an international environment" (p. 1). The TOEIC does not identify what an international environment is. Nor does it give any theoretical explanations of the aspects of English that the test measures or how the TOEIC version of English is different from world Englishes (Kachru & Nelson, 1996) or dialects of English that are encountered in natural, everyday life situations. Moreover, it is not clear whether the TOEIC's definition of English includes non-native speakers' (NNS) English. It is also ambiguous as to whether the TOEIC test developers took into account international communication between NNSs.

The TOEIC test makers do not show the construct validity of each section separately, although they do give their general concepts of the test. For example, the *TOEIC Technical Manual* ([www.toEIC.com](http://www.toEIC.com)) says, "The test does not require specialized knowledge or vocabulary beyond that of a person who uses English in everyday work activities" (l-1). Yet some of the settings and situations in the TOEIC such as those involving research, contracts, and negotiations must require cognitive academic language proficiency (CALP) in Cummins' (2001) sense. Furthermore, the TOEIC does not measure

examinees' pragmatic competence, which is included in most models of language proficiency and communicative competence (Bachman, 1990; Canale & Swain, 1980). Recent research on communicative competence and reading has come to view language and language proficiency in terms of context and purpose (Hudson, 1996; Savignon, 1991). The TOEIC does not appear to have a theoretical framework for its discrete multiple-choice format in Part 5 (Incomplete Sentences) and Part 6 (Error Recognition) of the reading section.

As Alderson (2000) states, in tests like the TOEIC, most commonly "the tester does not know why the candidate responded the way she did" (p. 212). For example, this is a sample question of Part 5 (*TOEIC Examinee Handbook*, 1998).

- Because the equipment is very delicate, it must be handled with \_\_\_\_\_.
- (A) caring
  - (B) careful
  - (C) care
  - (D) carefully. (p. 27)

In a real-world work environment, intelligibility seems to be important, particularly in international communication between NNSs. In an actual daily communication, answers B and D may not be critical errors with regard to intelligibility. Another example is also a sample question of Part 5 (*TOEIC Examinee Handbook*, 1998).

- Mr. Yang's trip will \_\_\_\_\_ him away from the office for ten days.
- (A) withdraw
  - (B) continue
  - (C) retain
  - (D) keep. (p. 28)

The test-takers should know an idiom *keep A from B* to answer this item. However, the question does not give any contexts related to the sentence. The form of Part 5 itself does not seem to correspond to real-world communications that test-takers may engage in. If the TOEIC examines only the test-takers' general English proficiency, the form of Part 5 without contexts can be acceptable. Yet the TOEIC developers contend that they measure English proficiency in an international work environment. In a daily communication, people can guess what interlocutors mean from contexts, which is different from the test form and test condition of Part 5 in the TOEIC.

For these reasons, test developers have to keep test-takers' responses to each item in mind when designing and choosing the items they intend to include in the TOEIC, as Alderson (2000) points out. Yet there are no available accounts of how this is done such as explained for the TOEFL (Peirce, 1992). Investigation by the test developers should consider the test-takers'

perspectives and attitudes toward the items and passages of the reading comprehension test. If research on construct validity from multicultural and multilingual perspectives has been already conducted, the results should be published and acceptable to the test-users, test-takers, and other language-testing researchers.

### *Content Validity*

A direct reading test should reflect as closely as possible the interaction that takes place between a reader and a text in the equivalent real life reading activity. However, in real life, reading purpose, background knowledge, formal knowledge, and various types of language knowledge may all interact with text content to contribute to a reader's text comprehension. (Weir, 1997, p. 39)

All these factors are crucial to the development of appropriate second language (SL) reading tests despite the difficulties they may entail. The Chauncey Group International, the developer of the TOEIC, seems to place importance on choosing test content that is consistent with the test-takers' needs. As explained in the *TOEIC Technical Manual*, studies (Tannenbaum & Rosenfeld, 1995<sup>2</sup>; Woodford, 1982, cited in the *TOEIC Technical Manual*) have assessed the English-language skills needed by employees from various parts of the world to meet the requirements of multinational companies. However, these studies have only focused on the English-language skills required by multinational companies, implying that the context for the test is corporate interest rather than the potential test-takers' real-world work context.

While reviewing the reading section of the TOEIC test (specifically Part 7 of the test), we realized that the test reflects no consideration of cultural and linguistic differences in examinees. Although the test claims to be "unbiased and culturally relevant for the test-takers worldwide" (*TOEIC Technical Manual*, II-2), the context and settings used as reading passages in the TOEIC test are clearly based on North American standards, which is an obvious disadvantage for others brought up in a different culture or applications of results of the test to situations outside North America. For example, this is a sample question of Part 7 (Reading Comprehension, *TOEIC Official Test Preparation Guide*, 2001).

Message for: Mr. Ibrahim

From: Michel leBlanc

Taken by: Henri

Time: 2:15 p.m., Thurs.

Message:

Michel LeBlanc at Batir Construction called. Has finished updating the contract but can't meet you on Friday at 3. Wanted to know when he

can reach you to reschedule. Will be at home this evening, but will try to contact you before then. If he doesn't get in touch with you, call him after 8 p.m. at home at 24-55-5123.

Sample Question 59

59. Why did Mr. LeBlanc call Mr. Ibrahim?

- (A) To rearrange a meeting
- (B) To ask for some building work to be done
- (C) To find out when a meeting will end
- (D) To request a work schedule

Sample Question 60

60. What is Mr. LeBlanc going to do?

- (A) Meet Mr. Ibrahim on Friday
- (B) Revise the contract
- (C) Go out for the evening
- (D) Telephone again this afternoon. (p. 185)

In this question, the test-takers cannot know in what kind of international environment this message is written. The test-takers do not know the relationship between Mr. Ibrahim and Henry in their company. What national company in what country Mr. Ibrahim and Henry are working for is not described in the sample either. Test-takers can guess nationalities and cultural backgrounds of people in the question by their names, but the information about the people is not clear. Despite lack of background information, these contexts seem to be important for natural communication in international workplaces, particularly in terms of pragmatics. For example, people call each other by their last name and position in Japanese and Venezuelan companies, whereas we are more familiar. If Henry is working in a company in those countries, he might have written his last name, and he might have written "Manager Ibrahim" on the message. The issue is not directly related to the answers of the questions (Questions 59 and 60), but the test-takers cannot have a clear image in what contexts this communication occurs when they answer the questions.

Another content-related issue in the TOEIC is the test developers' choice of using only one item format throughout the entire test. The problems of using multiple-choice questions for reading comprehension tests, as well as for other second-language tests, has been widely researched (Alderson, 2000; Weir, 1997). Three issues have often been discussed in the literature: (a) there is a guessing factor in which the examinee might be able to get an item right by a simple random elimination of the distractors; (b) the test-taker could also be able to answer by analyzing the structure of a question without really knowing the right answer; and most important (c) there is a chance that the examinees might be able to receive training in developing techniques for enhancing their ability to answer these types of questions correctly.

In addition, another problematic factor appears in the reading comprehension section of the TOEIC, specifically Part 6 (Error Recognition), in which the students have to identify the underlined word or phrase that is grammatically wrong or needs some kind of correction. We particularly agree with Enright et al.'s (2000) suggestion that isolated grammatical knowledge may not be critical for assessing reading comprehension. In fact, syntax-related features are also found in tests such as the TOEFL, but in a separate section specially designated for grammar structure and not as part of the reading comprehension construct as it is used in the TOEIC. We agree with Enright et al.'s perspective that the combination of sets and signals of a syntactic nature can enhance and provide support for other purposes such as adding contextual information or efficiently processing information.

### Limitations

As stated by many test-users and test-takers around the world (*TOEIC Examinee Handbook*), the TOEIC is not only recognized by important multinational companies and organizations, but it is also considered an effective tool that provides people with an accurate idea of a person's English proficiency level. The TOEIC seems to be practical and accurate, especially when used for screening purposes in the intermediate levels of English proficiency (Buck, 2001). As described by Buck, "although the test covers a wide ability range, the target group is probably best described as a lower intermediate level" (p. 210).

The test's validation, however, has a significant limitation. Although the TOEIC has established high internal reliability (e.g., the reading KR-Reliability Coefficient is 0.93), the construct and the content validity of the test have not been systematically and empirically examined since the test was developed in the 1970s.<sup>3</sup> Woodford's (1982) study has been the only published empirical study to examine the validation of the TOEIC that we know of since the test's development, and it reviewed only Part 7 of the Reading section. Richards (1992) suggests that more validation studies for the TOEIC are needed, but he does not examine its validation himself. Thus further studies of the test's validation are called for. Lack of construct and content validity research on the TOEIC may confuse test-users, test-takers, and English instructors. The test developers of the TOEIC do not show clear validations of the TOEIC such as what is standard English in an international environment and the process of test development in order to establish validity of the test (Bachman, 1990; Brown, 1995; Peirce, 1992).

### Conclusions

Approaches to language teaching and assessment have changed in many countries since the TOEIC was developed; most important, ideas about

validity in language testing have also developed over these two decades (Cumming, 1996; Kunnan, 1998, 2000). Yet only one systematic content validation study of the reading section of the TOEIC has been published since its development (Woodford, 1982).<sup>4</sup> Although we recognize the practicality and efficiency of the TOEIC, especially when used as a screening instrument, we would like to see more research done about the test's validity. Test developers should establish construct and content validity by illustrating the process of developing the test items in order to come up with a representation of the construct and the content according to the objectives the test was designed to measure.

As Richards (1992) points out, openness about the TOEIC content is crucial for investigating its validity. Bachman (1990) also points out "the consideration of test content is ... an important part of both test development and test use" (p. 244). The ethical standards of the test (see *Code of Fair Testing Practices in Education*, 1988) for test developers and test-users would be applied further as well in the process of demonstrating the TOEIC's validation. Bachman (1990) explains "demonstrating that a test is relevant to and covers a given area of content or ability is ... a necessary part of validation" (p. 244). Boyd and Davies (2002) claim "accountability in language testing ... requires openness to stakeholders" (p. 296). They also contend that openness to stakeholders is necessary for all test developers in order to establish an ethics of tests and that it is responsibility of test-makers to test-users and test-takers in regard to ethics of tests. Because the TOEIC has been adopted as a high-stakes test by many companies in the world, openness about its validity will facilitate establishment of ethics for the sake of its test-users and test-takers. We hope that the TOEIC developers have accountability for people who use this test as a high-stakes test all over the world.

### Notes

<sup>1</sup>Scoring of the TOEIC is described in the *TOEIC Technical Manual* on their Web site ([www.toeic.com](http://www.toeic.com)).

<sup>2</sup>The *TOEIC Technical Manual* does not include Tannenbaum and Rosenfeld's (1995) study in its references. We have been unable to locate this reference.

<sup>3</sup>Woodford's (1982) content validity study on the reading section is problematic because its sample size is small (99 participants) and participants were only Japanese. The study did not examine the content validity of Part 5 (Incomplete Sentences) and Part 6 (Error Recognition) as reading assessments.

<sup>4</sup>Most published validation studies of the TOEIC have examined the concurrent validity of the test (Wilson, 1993; *TOEIC Report on Test-Takers Worldwide*, 1996). See *TOEIC Technical Manual*, 1-2).

### Acknowledgments

We are grateful to Alister Cumming for his comments on earlier drafts of this article. We also thank the anonymous *TESL Canada Journal* reviewers and editors for their helpful comments.

Correspondence about this article should be addressed to Manami Suzuki at msuzuki@oise.utoronto.ca or MLC, OISE/UT, 252 Bloor Street West, Toronto, Ontario, Canada M5S 1V6.

### *The Authors*

Manami Suzuki is a doctoral student at the Ontario Institute for Studies in Education of the University of Toronto. Her academic research interest is in L2 writers' process of revision. She has been teaching English and language education as a lecturer at Dokkyo University and Tokyo Woman's Christian University in Japan since April 2004.

Carolina Daza is finishing her master's thesis and is interested in sociocultural theory and how peer interaction enhances second- or foreign-language learning.

### *References*

- Alderson, J.C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Boyd, K., & Davis, A. (2002). Doctors' orders for language testers: The origin and purpose of ethical codes. *Language Testing*, 19(3), 296-322.
- Brown, J.D. (1995). *Language testing in Japan*. Tokyo: Japan Association for Language Teaching.
- Buck, G. (2001). *Assessing listening*. Cambridge, UK: Cambridge University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Code of fair testing practices in education. (1988). Washington, DC: Joint Committee on Testing Practices. Retrieved from October 16, 2003, from <http://www.ericae.net/code.txt>
- Cumming, A. (1996). Introduction: The concept of validation in language testing. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 1-14). Clevedon, UK: Multilingual Matters.
- Cummins, J. (2001). *Negotiating identities* (2nd ed.). Ontario, CA: California Association for Bilingual Education.
- Enright, M.K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework: A working paper* (TOEFL Monograph Series MS-17). Princeton, NJ: Educational Testing Service.
- Gilfert, S. (1996). A Review of TOEIC. *Internet TESL Journal*, 2(8), 1-10. Retrieved October 16, 2003, from <http://iteslj.org/Articles/Gilfert-TOEIC.html>
- Hudson, T. (1996). *Assessing second language academic reading from a communicative competence perspective: Relevance for TOEFL 2000*. Princeton, NJ: Educational Testing Service.
- Kachru, B.B., & Nelson, C.L. (1996). World Englishes. In N.H. Hornberger (Ed.), *Sociolinguistics and language teaching* (pp. 71-102). Cambridge, UK: Cambridge University Press.
- Kim, H.J. (2001). Language learning strategies, learning styles and beliefs about language learning of Korean university students. *Journal of Pan-Pacific Association of Applied Linguistics*, 5(1), 31-46.
- Kunnan, A.J. (1998). *Validation in language assessment: Selected papers from the 17th language testing research colloquium, Long Beach*. Mahwah, NJ: Erlbaum.
- Kunnan, A.J. (2000). *Fairness and validation in language assessment: Selected papers from the 19th language testing research colloquium, Orlando, Florida*. Cambridge, UK: Cambridge University Press.
- Peirce, B.N. (1992). Demystifying the TOEFL reading test. *TESOL Quarterly*, 26, 665-689.
- Perkins, K. (1987). Test of English for international communication. In C.W. Stansfield (Ed.), *Review of English language proficiency tests* (pp. 81-83). Washington, DC: TESOL.
- Perkins, K. (1998). Assessing reading. *Annual Review of Applied Linguistics*, 18, 208-218.



- Richards, R.A. (1992). Review of the test English for International Communication. In J.C. Conoley (Ed.), *The eleventh mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Savignon, S. (1991). Communicative language teaching: State of the art. *TESOL Quarterly*, 25, 261-277.
- TOEIC *Examinee handbook*. (1998). Princeton, NJ. Educational Testing Service.
- TOEIC *Examinee handbook*. Retrieved October 16, 2003, from [http://www.toeic.com/2\\_4info/library.htm](http://www.toeic.com/2_4info/library.htm)
- TOEIC *official test-preparation guide*. (2001). Princeton, NJ: Educational Testing Service.
- TOEIC *report on test-takers worldwide 1997-98*. Retrieved October 16, 2003, from [http://www.toeic.com/2\\_4info/library.htm](http://www.toeic.com/2_4info/library.htm)
- TOEIC *technical manual*. Retrieved October 16, 2003, from [http://www.toeic.com/2\\_4info/library.htm](http://www.toeic.com/2_4info/library.htm)
- Weir, C.J. (1997). The testing of reading in a second language. In C. Clapham, & D. Corson (Eds.), *Language testing and assessment. Volume 7 of the encyclopedia of language and education* (pp. 39-51). Boston, MA: Kluwer Academic.
- Wilson, K. (1993). Relating TOEIC scores to oral proficiency interview ratings. *TOEIC Research Summaries, 1*. Princeton, NJ: Educational Testing Service.
- Woodford, P.E. (1982). The test of English for international communication (TOEIC). In C. Brumfit (Ed.), *English for international communication* (pp. 61-72). Oxford, UK: Pergamon Press.