

# Insights into Test of English for International Communication

Sachiko Nakamura

*Kyushu Sangyo University*

The test of English for International Communication (TOEIC) is an English-language proficiency test developed by Educational Testing Services (ETS). Since it was first administered in Japan in 1979, an increasing number of schools and companies have been adopting TOEIC scores as a measure of English ability for various purposes such as crediting, grading, and hiring. Accordingly, the need for students to obtain higher scores is on the rise. The aim of this paper is to provide an overview of the TOEIC test such as its history and score use as well as changes in the test format effective May 29, 2016. Followed by a discussion of reliability, validity, and washback of the TOEIC test, several pedagogical suggestions are offered.

TOEICとは、テスト開発機関(ETS)によって開発および制作された、英語コミュニケーション能力を測定する為の試験である。1979年に日本で初めて施行されて以来、単位認定、成績評価、雇用採用といった様々な目的で、多くの教育機関や企業に活用されている。本稿は、TOEICテストの歴史やスコアの活用法、また、2016年5月29日より導入される新問題形式に触れるとともに、試験の一貫性、信憑性、学習環境における影響に言及する。

The Test of English for International Communication (TOEIC) is an English-language proficiency test developed by Educational Testing Services (ETS). It “measures the everyday English skills of people working in an international environment. The scores indicate how well people can communicate in English with others in business, commerce, and industry” (ETS, 2012, p. 2). The test has become established as one of the predominant standardized English examinations in Japan and Korea especially, but also in many other countries throughout the world (ETS, 2013). In November, 2015, ETS announced a new format of questions in the TOEIC Listening and Reading test (hereinafter the TOEIC test), effective in May 2016. The aim of this paper is to examine the

traditional and newly-formatted TOEIC tests by utilizing sources from both inside and outside the company, i.e., ETS groups as well as foreign language or second language (L2) researchers and educators. First, an overview of the test is provided, followed by a description of the updates. Next, validity, reliability, and washback are discussed, and pedagogical suggestions are offered.

## **History of the Test**

The TOEIC test was first created in 1979 by Chauncey Group International Ltd., owned by the US-based Educational Testing Service (ETS), following a request from the Japanese Ministry of International Trade and Industry (MITI). The goal was to develop a test which could be used to evaluate English ability in business contexts. Focusing on international English, that is, the language used by non-native English speakers with other non-native English speakers, ETS sent a team of language specialists to Japan. They observed the English being employed by international business people in various settings in everyday workplaces. The test designers reflected the findings from the study and developed the TOEIC test, which is composed of Listening and Reading sections that “incorporate stimulus material found in the business world” (ETS, 2013, p. 2). Since December 2, 1979, when the test was first administered to 2,710 test-takers in Japan, and after it was revised in 2006, it has become “the world’s leading test of English-language proficiency” taken by approximately five million test-takers every year worldwide (ETS, 2013).

## **Current Status of the Test and its Score Use in Japan**

According to The Institute for International Business Communication (IIBC), approximately 2,400 corporations today use the TOEIC tests for hiring, screening, and promoting (IIBC, 2016). A survey conducted by IIBC in 2015 revealed that 60% of 267 companies consider TOEIC scores of job applicants as their skills and achievement in the process of screening (IIBC, 2016).

In another survey conducted by IIBC in 2015, approximately 45% of 1144 universities and colleges reported that they made use of TOEIC scores for

their entrance examinations, school credits, or both (IIBC, n.d). The English department at Sophia University, for instance, requires a score of 780 in order for candidate students to apply through *suisennyushi* (i.e., Entrance Test for Recommended Students) (Jouchi Daigaku Gakujikyoku Nyuugaku Senta, n.d.), while Meiji University gives six credits to students in the Economics department who have obtained a score of 840 or higher (Meiji University, 2014). Information regarding TOEIC score usage by different universities is available through this search engine <http://www.toeic.or.jp/toeic/about/data/search.html>. Based on a survey, IIBC (2013) reported that the average scores among undergraduate and graduate students in Japan are 562 and 599, respectively.

## Changes in the Test Format

In November 2015, ETS (2015) announced updates to the TOEIC test format. The main rationale for the changes is to focus on “more authentic communication” (p. 1) by means of adopting conversations among three people and formats reflecting communicative tools which have been frequently used over the last decade such as texting and online chatting. The new format also requires test-takers to utilize multiple sources simultaneously (e.g., listening to an announcement while looking at a map or chart) in order to answer questions. Other changes are inclusions of elisions (e.g., gonna), fragments (e.g., Could you?), and frequent exchanges of shorter utterances in spoken texts. Some tasks include questions at the discourse level (i.e., intentions of writers/speakers and implications) in both the Listening and Reading section. Table 1 summarizes and compares the current and new TOEIC test format. The scoring system, total test duration, and range of English accents (i.e., U.S., British, Canadian, Australian, and New Zealand) remain the same.

## Scoring

TOEIC scores are interpreted in terms of percentiles (i.e., norm-referenced test). Each test-taker’s answer sheet is read by an optical scanner and recorded by a software package called Integrated Operations, Processing, and Scoring (I-OPS) system. Scores are determined by the number of correct answers, and then

Table 1

*Format Comparison of the Current and New TOEIC Test*

The Current TOEIC Test		The New TOEIC Test		
Part 1	Photographs	10	Photographs	6
Part 2	Question – Response	30	Question – Response	25
Part 3	Conversations	30 (3x10)*	Conversations (with and without a visual image)	39 (3x13)*
Part 4	Talks	30 (3x10)*	Talks (with and without a visual image)	30 (3x10)*
Part 5	Incomplete Sentences	40	Incomplete Sentences	30
Part 6	Text Completion	12 (3x4)*	Text Completion	16 (4x4)*
Part 7	Reading Comprehension	0	Reading Comprehension	0
	Single Passages	28	Single Passages	29
	Double Passages	20	Multiple Passages	25

*Notes:* Numbers in parentheses describe the number of questions which are allocated in a set of question. For example, (3x10)\* means that there are 10 conversations, each of which is followed by three questions. Adapted from ETS (2015, p. 2).

converted into scaled scores through equating, a statistical procedure developed by ETS. The rationale for not using examinees' raw scores, according to ETS (2012), is to make the scores from different test administrations comparable, taking into account score variance because of inevitable differences in overall test difficulty. Examinees are graded on a scale of 5-495 on the Listening and Reading section, respectively, in increments of 5 points, for a combined maximum score of 990. They are also presented with percentile rank information based on a pool of test-takers from the past three years. The result of the test is posted to the test-takers in the form of an official score certificate, which consists of the information discussed above as well as a description of the English-language abilities typical of test-takers scoring at about the same level. A table of the descriptors can be found at [http://www.toEIC.or.jp/english/toEIC/guide04/guide04\\_02/score\\_descriptor.html](http://www.toEIC.or.jp/english/toEIC/guide04/guide04_02/score_descriptor.html).

## **Reliability**

“A test . . . should give the same results every time it measures” (Brown, 2005, p. 169). Test reliability functions as an indicator of the extent to which the results can be considered consistent or stable, and is generally estimated by calculating and interpreting reliability coefficients and standard error of measurement (SEM). ETS adopts internal-consistency reliability strategies and estimates the consistency of a test by analyzing the data from their norming samples, using the Kuder-Richardson formula (K-R20), a statistic procedure reported to provide the most accurate estimate (Brown, 2005, p. 185). ETS (2013) states that “the reliability of the TOEIC Listening and Reading section scores across all forms of our norming samples has been approximately 0.90” (p. 16), indicating high reliability. SEM is another statistical tool that can be used to evaluate reliability. Measurement error refers to differences in scores that are caused by extraneous sources rather than the purpose of the test. An examinee may be physically in a better (or worse) condition in one occasion than the other, or the test administration environment may be worse than the other times because of the noise from outside or the room temperature. SEM determines a band around a test-taker’s score within which his or her score would probably fall if the test were administered repeatedly (Brown, 2005). ETS (2013) reports that the SEM is approximately 25 scaled score points for each of the TOEIC Listening and Reading sections. This can be interpreted in the following way. Imagine that a test-taker scores 200 on the Reading section of the TOEIC test. The test-taker’s true score (as opposed to the obtained score) will fluctuate between 175 and 225, 68 percent of the time. While ETS (2015) claims that the new format holds the same test quality, no results of statistical data analyses have been yet publicized.

## **Validity**

High reliability alone is not sufficient enough to claim high test quality. In fact, it is possible for a test to be consistent without being valid (Bailey & Curtis, 2015; Brown, 2005). Validity, defined as “the degree to which a test measures what it claims, or purports, to be measuring” (Brown, 2005, p. 220), is another important criterion for evaluating a test. A study conducted by ETS in 2007

seems to provide the latest information about validity on the TOEIC test. ETS administered a self-report (can-do) inventory to approximately 5,000 TOEIC test-takers in Japan and Korea immediately after they had taken the test (Powers, Kim, & Weng, 2008). The inventory was composed of 25 reading tasks and 24 listening tasks in the form of can-do statements. Those statements were, for instance, “read the letters of the alphabet” (Powers, Kim, & Weng, 2008, p. 8) and “read and understand an agenda for a meeting” (p. 8) as reading tasks, as well as “understand simple questions in social situations” (p. 6) and “understand a co-worker discussing a simple problem that arose at work” (p. 6) as listening tasks. (For the full list of can-do statements, see Powers, Kim, & Weng, 2008). The participants were asked to indicate their ability to perform each task on a 5-point scale, ranging from 1 (*not at all*) to 5 (*easily*). ETS ranked each task by the level of its difficulty and grouped the test scores into seven levels. By means of calculating percentages of examinees who answered that they could perform the task either easily or with little difficulty at each of the seven levels, ETS examined the relationship between the two variables (i.e., test scores and self-assessments). Reporting a correlation of .57 between the listening tasks and TOEIC Listening scores, and .52 between the reading tasks and TOEIC Reading scores, ETS concluded that it is “a good evidence for the validity of the TOEIC scores” (Powers, 2010, p. 8).

Nevertheless, the validity of the TOEIC test seems to have been suffering. An examination of literature suggests that problems with the test validity by and large stem from the directionality and modality of the language skills addressed in the test. It has been argued that communicative proficiency cannot be measured in the written mode with multiple-choice items (Burrow, 2010; Sarich, 2014). Questions addressing sentence level comprehension rather than discourse level seem to be another element weakening the validity of the test purported to measure communicative language skills (Booth, 2012; Douglas, 2000). Booth (2012) points out that the test lacks important aspects of real-life communication such as indirect speech acts and pragmatic implication, as well as interactive language use such as natural hesitations and negotiations of meaning. Questions, then, arise as to to what extent the new version compensates for the

lack of authenticity of communicative language. As noted earlier, ETS seems to have addressed this issue by adopting exchanges consisting of more natural spoken and written texts as well as questions at the discourse level (for a sample of new item types, see ETS, 2015). However, further analyses of test items are vital to discuss the validity of the new TOEIC test.

## **Washback**

Washback is the effect, whether positive or negative, which a test creates on teaching and learning (Bailey & Curtis, 2015). ETS (2013) states that their products “promote learning and educational performance, and support education and professional development for all people worldwide” (p. 2). It also lists what the TOEIC test enables test-takers to do such as qualifying for a new position or promotion in a company, enhancing their professional credentials, and verifying their current level of English proficiency (ETS, 2012). In fact, many organizations are actively adopting the TOEIC test for these purposes. At the same time, the negative washback that the test has been creating on L2 teaching is notable. Despite the description of the TOEIC test by ETS, many language classrooms in universities and language institutions as well as commercially published textbooks are by and large underpinned by an approach focused on test-taking strategies. In this approach, the main resource is TOEIC-like materials, and students are taught various test-taking skills such as skimming, scanning, and inferencing. Memorizing business-related vocabulary and timed practice are also highly valued in this approach. While such instructions are merited, they may not be as effective in developing communicative language skills because they provide very few opportunities for students to interact with one another or use the language to communicate. ETS’s initial head TOEIC researcher, David Wood, warns that “studying past forms of the test may help the test taker become more familiar with the test item types, but will not help in actually improving a test score or a test taker’s overall proficiency” (Wood, 2010, p. 42). He added that merely memorizing TOEIC vocabulary will not generally improve a score if other aspects of L2 skills are not developed. It has also been pointed out that repeated practice with test questions tends to become

tedious and stifling, which results in discouraging students to learn (Bresnihan, 2013; Sarich, 2014). Unfair judgments caused by the misuse of TOEIC scores are another issue. Many language classes use TOEIC scores or gains in scores as a measure for grading students, which is not supported by norm-referenced tests. Despite the wide spread use of the TOEIC test, appropriate score interpretation does not seem to be well understood among many administrators and teachers. Negative washback caused by its business-oriented contents has also been reported by university professors teaching TOEIC preparation courses (Bresnihan, 2013; Yasunaga, 2014). In an EFL context such as Japan, students typically have limited exposure to the target language outside the classroom, let alone the experience at an international workplace. This lack of both content and formal schemata could cause struggle and poor performance on the test, regardless of learners' English proficiency itself (Bresnihan, 2013).

## **Pedagogical Considerations**

A discrepancy appears between the types of L2 skills that the TOEIC test is designed to measure and the types of skills that are typically focused on in TOEIC-related classrooms. Several suggestions can be offered to bridge the gap. One of them is to create various tasks focused on meaning and communication by means of modifying the TOEIC materials currently used in classrooms. For beginning students, for instance, pictures in Part 1 can be used for pair-work in which one tries to choose the correct picture based on the description provided by the other student. Using prepositions that commonly appear in Part 1, students play a timed game in which they describe the location of the objects in the classroom, taking turns while passing a pen. The student who is still holding the pen when the timer goes off is the loser.

For intermediate and advanced students, more complex tasks can be offered. For example, problem-solution is one of the most frequent genre features appearing in Part 3 (Yasunaga, 2014). Taken into an active classroom, students in pairs can create their own version of a telephone conversation between an online shopper complaining about the defective product delivered and a customer service representative offering a solution, and negotiate until they reach an



agreement. Afterward, the whole class can discuss what kinds of problems they created and the types of solutions offered. They can even vote for the worst problem or the best solution. These are only a few examples, and with a little creativity teachers can come up with further ways to develop a communicative classroom for TOEIC study (for further ideas about a communicative approach to teaching TOEIC-related courses, see, for example, Davies, 2005; Falout, 2005; Kim, 2010; Kudo, 2009).

## Conclusion

It is one thing to criticize the validity of the test and list its possible negative washback; however, it is clearly another to provide an effective methodology through which students will experience various ways of learning English while simultaneously developing communicative language competency and adequate skills to obtain high TOEIC scores. It is the author's hope that we as language teachers will facilitate such harmonious development of students' L2 proficiency.

## References

- Bailey, K. M., & Curtis, A. (2015). *Learning about language assessment: Dilemmas, decisions, and directions* (2nd ed.). Boston, MA: National Geographic Learning.
- Booth, D. K. (2012). Exploring the washback of the TOEIC in South Korea: A sociocultural perspective on student test activity. (Doctoral thesis, The University of Auckland). Retrieved from <https://researchspace.auckland.ac.nz/handle/2292/19379>
- Bresnihan, B. D. (2013). Useful information teachers and administrators should know about the TOEIC. In N. Sonda, & A. Krause (Eds.), *JALT2012 Conference Proceedings* (pp. 692-700). Tokyo, Japan: JALT.
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York, NY: McGraw Hill.
- Burrow, S. (2010-2011). TOEIC: The 100-hour recommendation and its implications. *Sapporo Daigaku Gaikokugo Gakubu Kiyou* [Language and Culture], 73, 13-26.

- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge, England: Cambridge University Press.
- Davies, A. (2005). A four-point approach to teaching TOEIC test preparation courses affectively and communicatively. *The Language Teacher, 29*(2), 9-13.
- ETS (Educational Testing Service). (2012). *Examinee handbook: Listening & reading*. Princeton, NJ: Author.
- ETS (Educational Testing Service). (2013). *User guide: Listening & reading*. Princeton, NJ: Author.
- ETS (Educational Testing Service). (2015). *Examples of new item types on the TOEIC test*. Tokyo, Japan: The Institute for International Business Communication.
- Falout, J. J. (2005). Focused tasks to proceduralize TOEIC learning strategies. In T. Newfields, Y. Ishida, M. Chapman, & M. Fujioka (Eds.), *Proceedings of the 2004 JALT Pan-SIG Conference* (pp. 38-44). Tokyo, Japan: JALT.
- IIBC (The Institute for International Business Communication). (2013). *TOEIC program: Data and analysis 2013*. Tokyo, Japan: Author.
- IIBC (The Institute for International Business Communication). (2016). *TOEIC program guide*. Tokyo, Japan: Author.
- IIBC (The Institute for International Business Communication). (n.d). *TOEIC listening & reading test nyuugaku shiken, tanni ninntei ni okeru katsuyou joukyou: Daigaku, tanki daigaku, koutou sennmonn gakkou* [Reports on the use of the TOEIC listening & reading test for entrance examinations and school credits: University, college, vocational school]. Retrieved from <http://www.toEIC.or.jp/toEIC/about/data/search.html>
- Jouchi Daigaku Gakujikyoku Nyuugaku Senta. (n.d). *2016 nenndo: Jouchi daigaku nyuugaku shiken gaiyou: Suisen nyuugaku shiken* [2016 school year: A summary of Sophia university's entrance examinations: Entrance tests for recommended students]. Tokyo, Japan: Author.
- Kim, J. (2010). Effective communicative language teaching in a test-preparation class: Is it possible? *Hawaii Pacific University TESOL Working Paper Series 8*(1, 2), 39-43.

- Kudo, T. (2009). Incorporating a communicative student-centered approach in a TOEIC preparation course. *Kinjou Gakuin Daigaku Ronshuu. Jimbun Kagaku-Hen*, 5(2), 47-61.
- Meiji University (2014, March 13). *Gaikokugo gaibu kenntei ninntei seido* [Foreign language examination system]. Retrieved from [https://www.meiji.ac.jp/seikei/govern\\_special/foreign.html](https://www.meiji.ac.jp/seikei/govern_special/foreign.html)
- Powers, D. E. (2010). *Validity: What does it mean for the TOEIC test?* Princeton, NJ: Educational Testing Service.
- Powers, D. E., Kim, H.-J., & Weng, V. Z. (2008). *The redesigned TOEIC (listening and reading) test: Relations to test-taker perceptions of proficiency in English*. Princeton, NJ: Educational Testing Service.
- Sarich, E. (2014). A guide to planning and executing a TOEIC preparation course. *The Language Teacher*, 38(1), 17-21.
- Yasunaga, A. (2014). Genre features of the listening texts in the TOEIC. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings*, pp. 699-707. Tokyo, Japan: JALT. Retrieved from [http://jalt-publications.org/files/pdf-article/jalt2013\\_070.pdf](http://jalt-publications.org/files/pdf-article/jalt2013_070.pdf)
- Wood, D. J. (2010). TOEIC materials and preparation questions: Interview with an ETS representative. *The Language Teacher*, 34(6), 41-45. Tokyo, Japan: JALT.

## Author bio

*Sachiko Nakamura* is a lecturer at Kyushu Sangyo University and holds an MA in TESOL from Anaheim University. She is currently working on an edited book on *Innovation in Japan*, with Hayo Reinders and Stephen Ryan, to be published by Palgrave Macmillan. [info@sachikonakamura.org](mailto:info@sachikonakamura.org)

Received: February 7, 2016

Accepted: June 22, 2016