

TOEIC Survey: Speaking vs. Listening and Reading

Masaya Kanzaki
Kanda University of
International Studies

Reference Data:

Kanzaki, M. (2015). TOEIC survey: Speaking vs. listening and reading. In P. Clements, A. Krause, & H. Brown (Eds.), *JALT2014 Conference Proceedings*. Tokyo: JALT.

Educational Testing Service (ETS), the developer of the Test of English for International Communication (TOEIC), launched a computer-based speaking test in 2006 to complement the multiple-choice TOEIC test, which measures only listening and reading abilities. Although the speaking test is still relatively unknown, it is gradually gaining ground in Japan, with the number of test-takers in 2013 reaching 14,700 (Institute for International Business Communication, 2014). This study compares the TOEIC speaking test scores of 87 university students with TOEIC listening and reading scores. First the correlations among the scores of the three tests were examined. Then the scores were analyzed to determine whether there were thresholds in listening and reading scores under which test-takers' speaking scores were uniformly low. In addition, a post-exam questionnaire was administered to survey what the participants thought about the speaking test in relation to the listening and reading tests.

Test of English for International Communication (TOEIC)の開発者であるEducational Testing Service (ETS)は、リスニング力とリーディング力しか測定しない選択式のTOEICを補完するため、コンピューター方式のスピーキングテストを2006年に開始した。スピーキングテストはまだそれほど知られていないが、日本における受験者は2013年に14,700人に達し(Institute for International Business Communication, 2014)、徐々に地位を確立しつつある。本研究では大学生87人のTOEICスピーキングテストのスコアとTOEICリスニング・リーディングテストのスコアを比較する。初めに3種類のテストのスコア間の相関関係を検証した。次にリスニングとリーディングのスコアがそれ以下だとスピーキングのスコアが一律に低くなる閾値が存在するのかわかるためにスコアの分析を行った。さらに、テスト後にアンケート調査を行い、受験者がリスニング・リーディングテストと比較して、スピーキングテストをどのように思ったか分析した。

THE TEST of English for International Communication (TOEIC), a multiple-choice test composed of listening and reading questions, became the most widely used English proficiency test in Japan in 2013 when the number of its yearly test-takers reached 2,361,000 (Institute for International Business Communication, 2014), surpassing the number of Test in Practical English Proficiency (EIKEN or STEP) test-takers (Eiken Foundation of Japan, 2014). Educational Testing Service (ETS), the developer of the TOEIC test, claim that “TOEIC test scores indicate how well people can communicate in English with others in the global workplace” (Educational Testing Service, 2013, p. 2). However, this assertion has been challenged, mainly because an English proficiency test that does not test speaking and writing skills cannot measure test-takers' communication abilities in English accurately. For example, Hirai (2008) compared TOEIC scores with STEP BULATS (STEP Business Language Testing Service) writing test scores and obtained a correlation of .46 among participants whose TOEIC scores were above 800, concluding that the correlation is not high

JALT2014 CONFERENCE PROCEEDINGS



enough for TOEIC scores to be a good indicator of business writing skills. Similarly, Hirai (2010) compared TOEIC scores with the STEP BULATS speaking tests and obtained a correlation of .66, which was not high enough to support ETS's claim that TOEIC is an indicator of communication abilities in the global workplace.

ETS publications used to report that there was a strong relationship between learners' TOEIC scores and their speaking abilities (e.g., Wilson, 1993; Woodford, 1982). However, after the addition of the computer-based speaking and writing tests to the TOEIC brand in 2006, ETS-commissioned studies started suggesting that speaking and writing abilities are better measured by the speaking and writing tests, respectively. For example, Liao, Qu, and Morgan (2010) compared scores of the four TOEIC tests obtained from 12,105 test-takers in Korea (79%), Japan (20%), and Taiwan (1%), and concluded that "each test measures distinct aspects of English language proficiency that cannot be adequately assessed by the other tests" (p. 11). Similarly, Liu and Constanzo (2013) examined the scores of the four tests obtained from 4,935 test-takers in Korea and found that "the listening, reading, speaking, and writing components of the TOEIC tests measure related but distinct English-language proficiencies" (p. 213).

Objectives of the present study

The TOEIC speaking and writing tests are still relatively unknown to the general public in Japan, and only a small number of learners have taken them. It appears that learners are reluctant to try the tests because they feel their English is not good enough, perhaps believing that they have to improve their listening and reading skills before taking the test. As the need to improve productive skills in English is now being increasingly stressed in various areas of Japanese society, it is worthwhile to examine these new additions to the TOEIC brand, although this study only looks at the speaking test. The writing test was not included because of time and budget constraints. The study was conducted to answer the following research questions:

1. How closely do the scores of the three types of TOEIC test correlate with each other?
2. Are there thresholds in the listening and reading scores under which test-takers' speaking scores were uniformly low, indicating the minimum level of listening and reading scores needed to perform adequately on the speaking test?
3. How did the test-takers find the TOEIC speaking test in relation to the listening and reading tests?

Method

Participants

Most of the participants in this study were taking TOEIC listening and reading preparation courses during the 2013 spring semester at a private university near Tokyo that specializes in foreign languages. They agreed to participate in the study in exchange for exemption from the 3,700 yen TOEIC test fee. Initially, 88 students took part but one of them was excluded from this study because his listening and reading test scores were far below his previous TOEIC scores.

The purposes of the study as well as the procedures and requirements were explained to the participants before they signed a consent form. It was explained that the results of the speaking test and the questionnaire would not affect their grades but that the results of the listening and reading tests would affect their grades in the course, as the TOEIC Institutional Program (TOEIC IP) test used in this study was one of six TOEIC tests used during the semester, the results of which accounted for a portion of the final grade.

Among the students, 77 were female and 10 male; 68 were in their 2nd academic year, 15 in their 3rd year, and four in their 4th year. Their average age was 19.8 years old. In terms of fields of study, there were 51 international communication majors, 22 English language majors, 12 international business majors, and two Chinese language majors. All the participants were native Japanese

speakers except for three native Chinese speakers, who were fluent in Japanese. Five of the study participants were enrolled in the TOEIC 860 course, another five in the TOEIC 730 course, 62 in the TOEIC 650 course and six in the TOEIC 600 course (the numbers indicate the target scores of the courses). The remaining six were not taking any TOEIC course.

Materials & Procedure

The three tests used in the study were administered as TOEIC IP, in which each institution sets the time, date, and place of the exam. The participants took the listening and reading tests on July 28, 2013 and the speaking test the following day. For the speaking test, they were divided into two groups of 44 because there were not enough computers for all participants to take the speaking test at the same time. Each group was divided into three subgroups, one with 20 participants, the other two with 12 each, according to the sizes of the rooms used for the speaking test. The results of the three TOEIC tests were provided by the Institute for International Business Communication, the administrator of the TOEIC in Japan. The scores of the three tests were compiled, and descriptive statistics as well as Pearson product-moment correlation coefficients were computed. The questionnaire was filled out in Japanese on a paper form, and the data was entered into a Microsoft Excel file for calculation and analysis. Descriptions of the TOEIC listening, reading, and speaking tests are provided in Appendix A.

Post-Speaking Test Questionnaire

A 9-item questionnaire was developed to explore the participants' opinions about the speaking test. The questionnaire asked

- whether the participants had previously taken the speaking test and, if yes, how many times they had taken it;
- whether they had prepared for the speaking test;

- how they had found the speaking test in relation to the reading and listening tests;
- which they liked better, the speaking test or the reading and listening tests;
- whether they had been at a loss what to do during the speaking test due to being unfamiliar with the exam format;
- how well they had managed to demonstrate their speaking ability in the speaking test;
- whether they would like to take the speaking test again; and
- whether they thought their speaking score would improve if they took it again.

After filling in the questionnaire, the participants were encouraged to write comments on the three TOEIC tests. The questionnaire was administered after the speaking test on July 29, 2013 on campus. Eighty-three participants completed the questionnaire. The questions can be found in Appendix B (Japanese) and Appendix C (English translation).

Results

Descriptive Statistics

The scores of the three TOEIC tests were analyzed using SPSS (IBM Corp., 2013), and the results of the questionnaire were compiled.

Table 1 shows the descriptive statistics for the three tests. Overall, the participants' results were better for listening than reading. The average listening score was 105.41 points higher than the average reading score.

Table 1. Descriptive Statistics for Three TOEIC Tests (N = 87)

Test (possible score)	M	SD	Minimum	Maximum
TL (5-495)	370.98	54.06	205	495
TR (5-495)	265.57	65.55	115	400
TCLR (10-990)	636.55	109.86	320	890
TS (0-200)	113.45	25.69	40	160

Note. TL = TOEIC listening test; TR = TOEIC reading test; TCLR = TOEIC listening and reading tests combined; TS = TOEIC speaking test.

Correlations

The correlations between the scores of the three TOEIC tests were computed (see Table 2). Among the three combinations, the highest correlation is between the listening and reading scores ($r = .68$), and the lowest is between the reading and speaking scores ($r = .48$). This order is in agreement with the correlations reported by Liao et al. (2010): .76 between the reading and listening tests, .66 between the listening and speaking tests, and .57 between the reading and speaking tests. The correlations are also in agreement with those reported by Liu and Costanzo (2013): .73, .63, and .54 in the same order.

Table 2. Correlations Between Three TOEIC Tests (N = 87)

Test	TL	TR	TS
TL	1	.68*	.50*
TR		1	.48*
TS			1

Note. TL = TOEIC listening test; TR = TOEIC reading test; TS = TOEIC speaking test; * = $p < .001$.

The correlation between the speaking scores and the combined scores of the listening and reading tests was .53 ($p < .001$).

Scatter Plot Results

Three sets of scatter plots were created in order to determine whether there is a threshold for the listening score, the reading score, or the total score of the listening and reading tests, under which the participants' speaking scores are uniformly low. Figure 1 shows the relationship between the speaking and listening scores, Figure 2 between the speaking and reading scores, and Figure 3 between the speaking score and the listening and reading scores combined.

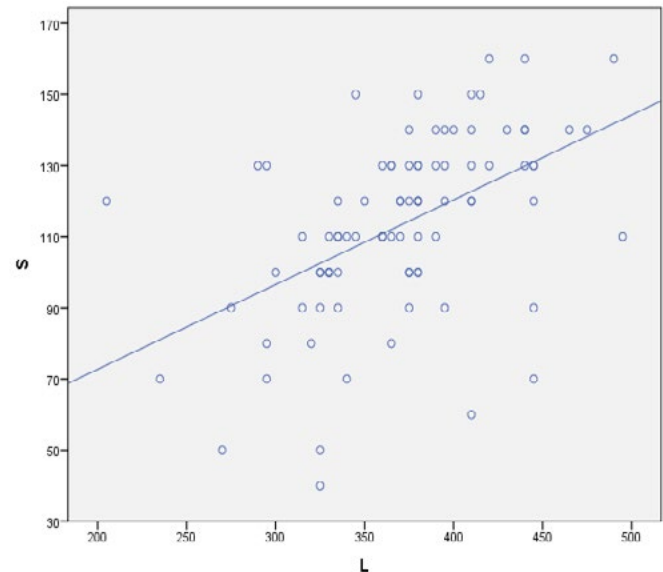


Figure 1. The relationship between the speaking scores and the listening scores. S = speaking score; L = listening score.

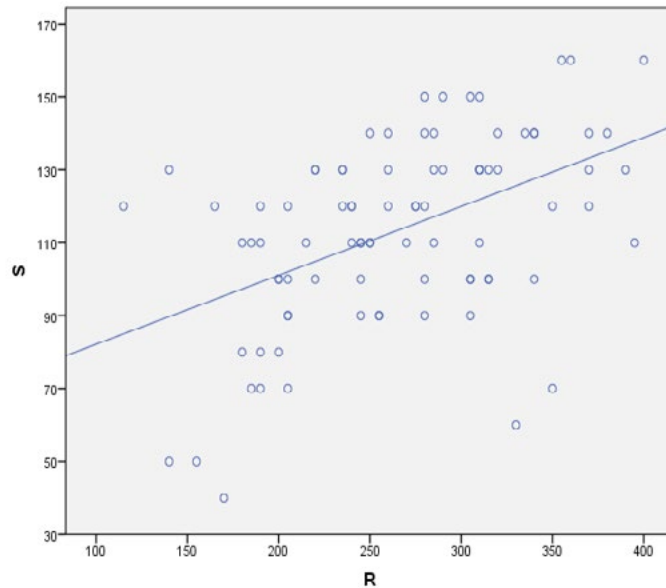


Figure 2. The relationship between the speaking scores and the reading scores. S = speaking score; R = reading score.

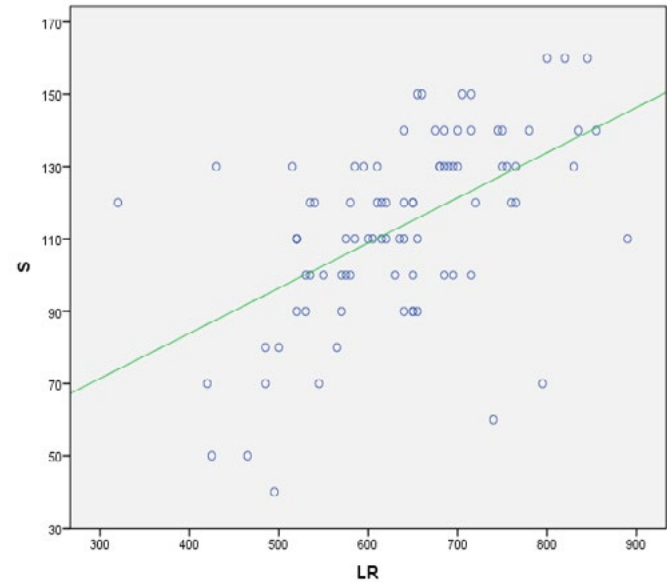


Figure 3. The relationship between the speaking scores and the reading and listening scores combined. S = speaking score; LR = reading and listening scores combined.

No threshold was detected in any of the three sets of scatter plots, as some students who received lower-than-the-average listening and reading scores did relatively well on the speaking test.

Post-Speaking Test Questionnaire Results

In order to explore what the participants thought about the speaking test, a questionnaire was conducted the day after the test, and was completed by 83 out of 87 participants. Only three respondents

indicated that they had previously taken the speaking test, but 72 indicated that they had tried to familiarize themselves with the test format by reading an informational booklet, visiting a website, or attending a workshop on it. Therefore, the majority of the participants had some degree of knowledge about the speaking test before taking it.

In question 3, the results of which are shown in Table 3, participants were asked how they had found the speaking test in relation to the reading and listening tests. Multiple answers were allowed. Of the 83 respondents, 54 indicated they had found the speaking test difficult in terms of question difficulty and 27 felt the test was tough in terms of concentration. Even so, 23 indicated that taking the speaking test was fun.

Table 3. Student Responses to “How Did You Find the TS in Relation to the TL and TR?” (N = 83)*

No. responses	In terms of questions		In terms of concentration			
	Easy	Difficult	Easy	Tough	Fun	Boring
	2	54	2	27	23	0

Note. TL = TOEIC listening test; TR = TOEIC reading test; TS = TOEIC speaking test; *Multiple answers were allowed.

In question 4 (see Table 4) participants were asked which they liked better, the speaking test or the reading and listening tests. The number of respondents who indicated that they preferred the speaking test is slightly higher than the number who preferred the reading and listening tests.

Table 4. Student Responses to “Which Do You Like Better, the TS or TL and TR?” (N = 83)

	I like TS.	I like TL & TR.	I like both equally.	I don't like either.
No. responses	32	30	7	14

Note. TL = TOEIC listening test; TR = TOEIC reading test; TS = TOEIC speaking test

In question 5 (see Table 5) participants were asked whether they were at a loss what to do during the speaking test due to being unfamiliar with the test format. Seventy-seven respondents indicated that they were at a loss, which suggests that they did not perform at their best in the test. That was reflected in their answers to Question 6, in which they were asked to indicate what percentage of their speaking ability they believed had been demonstrated during the test. The average was 41.38%: the highest being 100% and the lowest 1%. The number of responses to Question 6 was 81.

Table 5. Student Responses to “Were You at a Loss during the Speaking Test Due to Being Unfamiliar with its Format?” (N = 83)

	Many times	A few times	Not at all
No. responses	41	36	6

For Question 7, participants were asked whether they wanted to take the speaking test again. Sixty-one respondents indicated that they did. The results for question 7 are shown in Table 6.

Table 6. Student Responses to “Do You Want to Take the Speaking Test Again?” (N = 83)

	Yes, very much so	Yes, I do	I'm not sure	Not very much	Not at all
No. responses	28	33	12	6	4

In Question 8 (see Table 7) participants were asked whether they thought their scores would improve if they took the speaking test again. This question was intended to gauge the appropriateness of the difficulty level of the test on the assumption that if the test-takers' abilities were far below the level needed to perform the test tasks adequately, they would not feel they could perform better next time. On the other hand, if their abilities were slightly below the level, they would feel they could improve their scores next time. Fifty-four respondents thought that their scores would improve next time.

Table 7. Student Responses to “Do You Think Your Score Will Improve if You Take the Speaking Test Again?” (N = 83)

	Yes, very much so	Yes, I do	I'm not sure	Not very much	Not at all
No. responses	18	36	19	9	1

Forty-six respondents added comments at the end of the questionnaire. Eleven mentioned that the main problem in responding to the questions in the speaking test was that they did not have enough time to prepare their responses; they had to think quickly in order to answer the questions within the allotted time. Six of them expressed negative feelings towards the listening and reading tests, stating that achieving high scores in the TOEIC listening and reading tests does not mean having a good command of English.

Three students mentioned that they were aware of the presence of the other test-takers in the room, which made them feel hesitant to speak English. Two compared the speaking test with other English proficiency tests that have speaking components. One of them had taken the TOEFL iBT and mentioned there were similarities between the two tests. The other student mentioned that a speaking test should have a face-to-face interview format like the one used by IELTS because there is communication between the interviewer and the interviewee, implying that the computer-based TOEIC speaking test does not measure communication skills. Four mentioned that they regretted that they had not done enough preparation for the speaking test. Three students stated that the speaking test should be used more widely. Four compared the speaking test with the reading and listening tests in terms of the length, and they preferred the speaking test because it was shorter, meaning they could keep their concentration during the test. Other comments included “I was relaxed because I didn't have to worry about the score,” “It was a good test to measure my speaking ability,” “The speaking test requires a different type of concentration from the reading and listening tests and I felt test-taking strategies are also needed for the speaking test,” and “It was difficult to express my thoughts verbally.”

Discussion

The results showed that the scores of the speaking test correlated at .50 with the scores of the listening test, at .48 with the scores of the reading test, and at .53 with the reading and listening scores combined, and these correlations are statistically significant at the $p = .001$ level. However, correlations around .50 are not high enough for the reading and listening scores to be a predictor of learners' speaking abilities. According to Dorans (1999), the correlation between the predictor and the score to be predicted should be above .866 in order to achieve 50% certainty. These relatively weak correlations could be attributed to the fact that the participants were not familiar with the speaking test. The post-exam questionnaire

revealed that more than 90% of the respondents were at a loss what to do during the speaking test because they were not familiar with the test format, even though the majority of them had made efforts to learn about it, and they felt that they had not fully demonstrated their speaking abilities in the test.

In this study I attempted to find listening and reading score thresholds under which test-takers' speaking scores remain uniformly low. However, no such thresholds were found in the test results of the participants, whose average TOEIC listening and reading score was 636.55, with the lowest being 320 and the highest 890. One possibility for this is that such thresholds do not exist. Another possibility is that a threshold may exist but everyone in this group was above the threshold and thus it was not detected. If this was the case, it might be possible to find a threshold by investigating a group with lower proficiency.

The questionnaire answers showed that the participants felt positive towards the speaking test. Even though more than 60% of the respondents indicated that the speaking test was difficult, about a quarter of them found it fun and 70% indicated that they wanted to take it again. However, the participants of this study were language majors and most of them were interested in improving their spoken English. If the same questionnaire had been administered to others who are not interested in improving their spoken English, the results may have been different.

Conclusion

The correlations between the three tests were not strong enough for the listening and reading scores to serve as a valid predictor of learners' speaking abilities. It would be better to use the speaking test than the reading and listening tests when participants' speaking skills are to be assessed.

No thresholds in the listening and reading scores under which participants' speaking scores were equally low were found in this

study. The lowest listening and reading score among the participants was 320. Therefore, learners whose combined reading and listening scores are higher than 320 should not feel that their skills are too low to try the speaking test.

The questionnaire survey revealed that the speaking test was well received by the participants. With 70% of the respondents indicating that they wanted to take it again, the speaking test would be a better motivation booster than the reading and listening tests for those who are interested in improving their speaking skills.

Acknowledgments

The author appreciates the help of Dr. Yuki Shibuya of Kanda University of International Studies in developing the questionnaire used in this study and performing statistical analyses. This study was supported by JSPS KAKENHI Grant Number 25370727.

Bio Data

Masaya Kanzaki teaches at Kanda University of International Studies. His research interests include vocabulary acquisition, language testing, and corpus linguistics. <kanzaki-m@kanda.kuis.ac.jp>

References

- Dorans, N. J. (1999). Correspondences between ACT and SAT I scores. *College Board Report*, 99(1). New York: College Entrance Examination Board.
- Educational Testing Service. (2013). *TOEIC user guide—listening and reading*. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Tests/Test_of_English_for_International_Communication/TOEIC_User_Gd.pdf
- Eiken Foundation of Japan. (2014). *2013 Nendo juken joukyou gakuseki betsu* [Test-taking situations according to types of schools that test-takers attended in 2013]. Retrieved from http://www.eiken.or.jp/eiken/merit/situation/pdf/student_2013_03.pdf

- Hirai, M. (2008). Correlation between STEP BULATS writing and TOEIC® scores. In T. Newfields, P. Wanner, & M. Kawate-Mierzejewska (Eds.), *Proceedings of the 7th Annual JALT Pan-SIG Conference* (pp. 36-46). Tokyo: JALT. Retrieved from <https://jalt.org/pansig/2008/HTML/Hirai1.htm>
- Hirai, M. (2010). Correlation between STEP BULATS speaking and TOEIC® scores. In E. Skier & T. Newfields (Eds.), *Proceedings of the 8th Annual JALT Pan-SIG Conference* (pp. 12-25). Tokyo: JALT. Retrieved from <http://jalt.org/pansig/2009/HTML/Hirai.htm>
- IBM Corp. (2013). *IBM SPSS Statistics for Windows, Version 22.0* [Computer program]. Armonk, New York: IBM.
- Institute for International Business Communication. (2014). *TOEIC Tesuto Jyukennsyano Suii* [The changes in the numbers of yearly test-takers of the TOEIC Listening and Reading Test]. Retrieved from http://www.toEIC.or.jp/library/toEIC_data/toEIC/pdf/about/transition.pdf
- Liao, C., Qu, Y., & Morgan, R. (2010). The relationships of test scores measured by the TOEIC listening and reading test and TOEIC speaking and writing tests. *TOEIC compendium 2*. Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/Media/Research/pdf/TC-10-13.pdf>
- Liu, J., & Costanzo, K. (2013). The relationship among TOEIC listening, reading, speaking, and writing skills. *TOEIC compendium 2*. Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/Media/Research/pdf/TC2-02.pdf>
- Wilson, K. (1993). Relating TOEIC scores to oral proficiency interview ratings. *TOEIC research summary number 1*. Princeton, NJ: Educational Testing Services. Retrieved from <https://www.ets.org/Media/Research/pdf/TOEIC-RS-01.pdf>
- Woodford, P. E. (1982). An introduction to TOEIC: The initial validity study. *TOEIC research summary*. Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/Media/Research/pdf/TOEIC-RS-00.pdf>

Appendix A

Descriptions of TOEIC Listening, Reading and Speaking Tests

ETS treats the listening and reading as two sections of one test. The two tests are always administered together, and the scores of the two tests are added together in the score report. The speaking test, on the other hand, is treated as an independent test, and even when it is taken with the writing test, the speaking score and the writing score are not added together.

The listening test consists of 100 multiple-choice questions, and raw scores of between 0 and 100 are converted to scaled scores of between 5 and 495. The test has four parts, the details of which are shown in Figure A1.

Part	Task	# of Qs
1	For each question with a photo, listen to four sentences and choose the one that best describes the image.	10
2	Listen to a question or statement followed by three responses and choose the most appropriate response.	30
3	Listen to a conversation and answer comprehension questions.	30
4	Listen to a short talk and answer comprehension questions.	30

Figure A1. The four parts of the listening test.

The reading test consists of 100 multiple-choice questions, and raw scores of between 0 and 100 are converted to scaled scores of between 5 and 495. The test has three parts, the details of which are shown in Figure A2.

Part	Task	# of Qs
5	Choose a word or phrase to fill in a blank in a sentence.	40
6	Choose words or phrases to fill in blanks in a passage.	12
7	Read a passage or a set of two passages and answer comprehension questions.	48

Figure A2. The three parts of the reading test. The reading test starts with Part 5 because it immediately follows the listening test, which ends with Part 4, and the two tests are always taken as a set.

The speaking test is computer-based and requires test-takers to sit in front of a computer wearing a headset with a microphone. Instructions are provided on the computer screen and through the headset. Test-takers speak into the microphone, and what they say is recorded and sent to certified raters for evaluation. There are 11 questions, and scores are given in the range of 0 to 200. Figure A3 shows the details of the test.

Question	Task
1-2	Read aloud the text that appears on the screen.
3	Describe the picture on the screen.
4-6	Answer three questions about a single topic as though you are participating in a telephone interview.
7-9	Read the information on the screen and answer three questions about it as though you are responding to a telephone inquiry.
10	Listen to a recorded message about a problem and propose a solution for it.
11	Express an opinion about a specific topic.

Figure A3. Details of the speaking test.

Appendix B

Post-Speaking Test Questionnaire (Japanese)

- Q1. 以前TOEICスピーキングテストを受けたことがありますか?
 1) 今回が初めて 2) 以前受けたことがある(回)
- Q2. TOEICスピーキングテストに対する準備はしましたか?(複数回答可)
 1) TOEIC SW Test Guide(説明会で配ったブックレット)を見て問題形式の確認をした。
 2) TOEIC SWのサイト上で問題形式の確認をした。
 3) TOEICスピーキングテストの対策書を使って学習した。
 4) TOEICスピーキングテストのワークショップに出た(上原先生、横川先生、その他)。
 5) 何もしなかった。
- Q3. TOEICスピーキングテストをTOEICリスニング・リーディングテスト(普通のTOEIC)と較べて、どう感じましたか?(複数回答可)
 1) 易しかった 2) 難しかった
 3) 楽だった 4) 大変だった
 5) 楽しかった 6) つまらなかった
- Q4. TOEICスピーキングテストとTOEICリスニング・リーディングテスト(普通のTOEIC)とどちらが好きですか?
 1) TOEICスピーキングテスト
 2) TOEICリスニング・リーディングテスト(普通のTOEIC)
 3) 両方好き 4) 両方嫌い
- Q5. TOEICスピーキングテストの問題形式に慣れていないため、試験中、戸惑うことはありましたか?
 1) かなりあった 2) 少しあった
 3) まったくなかった
- Q6. TOEICスピーキングテストで自分のスピーキングの実力を何パーセント出せたと思いますか?
 _____ %

