# Preliminary Research on the Implementation of Computer-Based Speaking and Writing Test and its Implications

Terumi Miyazoe
Tokyo Denki University, Japan
t.miyazoe@mail.dendai.ac.jp

**Abstract**: This paper is a preliminary research on the implementation of an official computer-based Test of English for International Communication (TOEIC) speaking and writing (SW), a relatively new concept in the country. With the growing demand for graduates to possess competence in communication, the objective measurement of English proficiency has become vital in higher education. Contrary to the receptive skills of listening and reading (LR), a systematic evaluation of productive skills, such as SW, lags behind in terms of implementation. Such situation is evident among students with intermediate to low proficiency. This research was conducted among students in a science university whose major was not language. Data from two sources, namely, the test scores and the post-test questionnaire of examinees, were examined. Results in SW were analyzed with their most recent TOEIC LR results. Similar to prior studies (Liao, Qu, & Morgan, 2010), this paper obtained moderate-level correlation among the four skills, indicating that students in the LR-centered curriculum exhibit adequate development in SW proficiency. Moreover, the students regarded the computer-based SW test as fun. Overall, the computer-based SW test is a viable option for the objective assessment of the productive skills of students with intermediate and false-beginner proficiencies.

***Keywords:*** *Computer-based Test, TOEIC Speaking & Writing, Learner Perceptions, Correlational Analysis, International Competence, Curriculum Development*

## BACKGROUND

This research was conducted to provide basic information on the implementation of a computerized test for English speaking and writing (SW) among science-major students in a university in Japan.

Higher education faces a growing demand to produce students with specialized academic knowledge and skills, as well as a high level of communicative competence to work with people from around the globe. Thus, demands for an objective and systematic evaluation of productive skills, such as SW, in language education curriculum are rapidly increasing. However, a test that is cost-effective yet fair and accurate in processing evaluation of a large number of students is not easy to find.

Similar to the Test of English as a Foreign Language Internet-Based Test (TOEFL iBT, n.d.), the Test of English for International Communication SW (TOEIC SW) is particular in two ways. Following the English instructions on a personal computer (PC), which is connected to the test agency network, the examinees are required to:

1. talk to the PC to record their oral responses; and
2. type their writings by using the keyboard.

Therefore, in addition to the targeted SW ability, the test demands for a certain level of information and communication technology (ICT) skills. This test condition differs from another internationally recognized four-skill test, such as the International English Language Testing System (IELTS, n.d.), which is held face-to-face and with pen-and-paper. However, IELTS is much more costly because of the involvement of human examiners.

The full scores for TOEIC listening and reading (TOEIC LR) is 990 (L: 495, R: 495) and SW is 400 (S: 200, W: 200). The test measures the examinees' English proficiency for international communication, and their competency levels from false-beginner to advanced semi-native. Different from the LR test, which should be taken as a set of listening and reading, each SW test can either be taken independently or as a set, depending on the needs of the examinees; thus, test fees are reduced accordingly.

Prior to the current study, the English Testing Service (ETS), which provides TOEIC as well as other standardized tests, such as GRE and TOEFL in the United States, conducted a correlational analysis on the TOEIC four-skill scores between the Korean and the Japanese populations (Liao et al., 2010). The computer-based oral and aural training of university students were likewise tested, and the study concluded the high-level readiness of students for computer-based communication and learning (Miyazoe & Anderson, 2010; Miyazoe & Anderson, 2012).

## RESEARCH QUESTIONS

No prior research has implemented TOEIC SW at the university sampled in this study. The following questions are explored:

1. To what extent is the test functional for students with varying levels of English proficiency?
2. How do the students react to the computerized SW test?
3. To what extent are the obtained results comparable to the general data with a larger population outside?

The researcher was uncertain regarding the implementation of this experiment because majority of the students involved have not received structured/intensive instruction to develop their SW skills. On one hand, they may exhibit difficulty in adapting to the computer-based test environment and deliver a lower performance than they would in other test methods. On the other hand, their readiness in ICT, fostered by the science university, may allow for a positive new experience of computer-based testing.

## METHODOLOGY

### Implementation

This research was made possible with the propeller test program organized by The Institute for International Business Communication (IiBC), a branch agency of the ETS in Japan that accorded 20 examinees taking the TOEIC SW test for free. Normally, each test costs approximately USD 100 or twice more expensive than its LR version.

The test was conducted in a computer laboratory. A day before the test, the IiBC set up the necessary equipment. A total of 20 notebook PCs were used and connected online with the network of the test agency.

Prior to the test day, the students were provided with a few materials and methods to familiarize themselves with the test: e-mail of information on web links regarding the test formats (IiBC, n.d.), books containing practice tests, and face-to-face assistance during the practice test. The students accessed at least one of the aforementioned means to be oriented with the particular test format prior to the test day. All students claimed to take the TOEIC SW for the first time.

**Participants**

This study included 18 student participants, of whom 17 test data were analyzed. The students belong to the Engineering department of a science university in Tokyo. This middle-sized university includes approximately 10,000 enrollees. The participants varied in age and field of study, with 13 males and 4 females from first to junior year. A returnee who was born and spent his childhood in the US for six years was included. All students voluntarily participated in the exam. Given that the school population comprises 6% female students (Obunsha, n.d.), the percentage of female students, that is, 23.5%, is higher based on the school's general profile. Three students were enrolled in the night school program (the university offers day and night programs to cater to students working in science-related fields), which accepts students of various backgrounds, including adult working students.

**Data sources**

Three sources of data were referenced for the analysis:

1. TOEIC SW scores
2. TOEIC LR scores
3. Post-questionnaire

With the written consent of participants for possible publication, the TOEIC SW and LR scores of the 17 students were analyzed. The highest LR scores were included for students who have taken the LR test within a year of the current SW test day (many had taken the test few months prior to the current SW in the same semester).

The IiBC administered the post-questionnaire, which contained 16 questions to determine the perceptions of examinees regarding their SW test experience as well as their demographic data, such as age, gender, experience of staying abroad, and field of study, among others. A separate consent for analysis and possible publication of the post-questionnaire results was earlier obtained from the students and the agency, under the condition of anonymity and subsequent sharing of the research results with the participants.

Data were analyzed using EXCEL and SPSS. Only the results that show suggestive features are reported.

## RESULTS

**General LR/SW score description**

Table 1 and Figure 1 present the descriptive analysis of scores in the four skills.

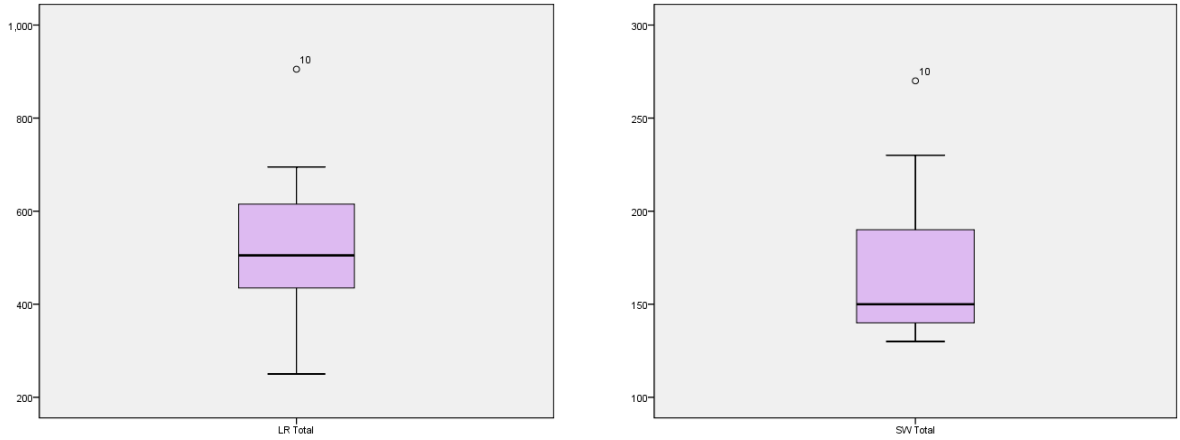
**Table 1.** Descriptive analysis of TOEIC scores in the four skills

| | N | Min. | Max. | Mean | SD |
|---|---|---|---|---|---|
| S | 17 | 50 | 130 | 74.1 | 23.2 |
| W | 17 | 60 | 140 | 95.3 | 24.3 |
| SW Total | 17 | 130 | 270 | 169.4 | 41.9 |
| L | 17 | 135 | 495 | 295.9 | 78.0 |
| R | 17 | 115 | 410 | 225.6 | 81.4 |
| LR Total | 17 | 250 | 905 | 521.5 | 151.3 |

The lack of prior data for TOEIC SW at the current study area renders it impossible to provide a general argument regarding the students of the sample university. However, in terms of mean scores, based on the Proficiency Scale (IiBC, 2012a) modified from a research on test takers' Can-Do perceptions and the TOEIC test score distribution (Powers, Kim, Yu, Weng, & VanWinkle, 2009), the average LR score of 521.5 corresponds to band C (between 470 and 730). This result indicates that the students possess communicative competence to cope with daily needs and business under limited conditions.

The two boxplots of Figure 1 confirm one outlier for each of LR and SW scores, which was from the returnee student. The average score of LR without the outlier is 498; thus, the outlier score is not considered as significantly influential in this study. The average score in TOEIC LR of approximately 500 is fairly representative of the general profile of the university students in Japan (IiBC, 2012c). However, the sample size is small given its exploratory nature, which validates this experiment as a model for further practice in most universities in Japan.

By contrast, the conversion table from previous data provided by IiBC (2012b) estimates that LR 500 corresponds to approximately SW 210 (S: 100, W: 110). Therefore, despite the estimation, 169.4 in TOEIC SW (163 without the outlier) in the current study may still be improved if the students are more familiar with and re-take the test.




TOEIC LR Total TOEIC SW Total

**Figure 1**. Score distribution of TOEIC LR and SW

Alternatively, the SW scores in this study were lower than expected because the curriculum lacks measures to foster the productive skills of SW at a similar level to those of LR.

**Gender differences**

Table 2 and Figure 2 summarize the results of the scores in the four skills according to gender. The current sample size was limited considering the exploratory nature of this study. Thus, when comparing LR and SW scores, the female group achieved higher scores in SW (180.0 vs. 166.2) and lower scores in LR (508.8 vs. 525.4) as compared with the male group.

Earlier data on SW scores provided by IiBC allow further interpretation (IiBC, 2012b). The average SW scores in 2011 official tests (sample size of almost 4,200) and IP tests (almost 5,000) indicate that the female group achieved higher scores in SW, with the range from 3.5 to 15.3 in SW total scores for both sub-groups of employees and students. Therefore, compared with males, female students are likely to score higher in productive skills.

**Table 2.** Descriptive analysis of TOEIC scores in the four skills according to gender

| | | S | W | L | R | SW Total | LR Total |
|---|---|---|---|---|---|---|---|
| Males | Mean | 70.0 | 96.2 | 296.5 | 228.9 | 166.2 | 525.4 |
| | SD | 21.2 | 25.3 | 87.1 | 89.0 | 41.5 | 168.7 |
| Females | Mean | 87.5 | 92.5 | 293.8 | 215.0 | 180.0 | 508.8 |
| | SD | 27.5 | 23.6 | 46.4 | 58.5 | 47.6 | 89.0 |

The balance in scores between LR and SW may not be proportional, but may be gender-specific. Curricula oriented to receptive skills (that is, as experienced by the majority in Japan, possibly due to cost issues) could pose disadvantages for female students compared with male students in terms of decision-making and evaluation.

Figure 2 presents the distribution and outlier effects in SW scores by gender. The difference in SW total scores between male and female groups was due to the lower score obtained by the male students, particularly in speaking. The difference in mean scores becomes more significant when two outlier and extreme point cases are excluded (S: 62.7; W: 88.2; SW total: 150.9). Whether this result was due to gender difference, the computer-based TOEIC SW test configuration, or both, was not evident. Nevertheless, although a curriculum does not intensify productive skills, the students may maximize the instructions by focusing on receptive skills, and develop the corresponding productive skills through knowledge and skill inference.

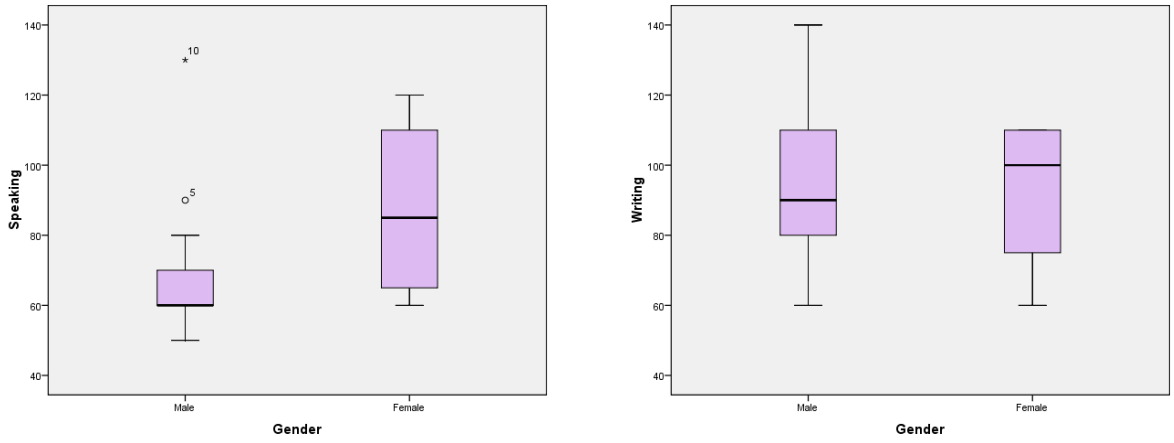



**Figure 2**. TOEIC SW score distribution according to gender

Alternatively, the TOEIC tests, or English proficiency tests in general, contain features that would benefit females in terms of SW skills and males for LR skills. Such features could be the underlying factors that might have led to the gender differences observed in previous research and in the current study.

**Perceptions towards TOEIC SW**

The post-test questionnaire administered by IiBC, as a support service in the propeller test program, has been relevant in providing a third-party perspective on the experimentation. A graphically processed analysis of the SW test results was obtained. With IiBC consent, Figure 2 presents data from the post-questionnaire.

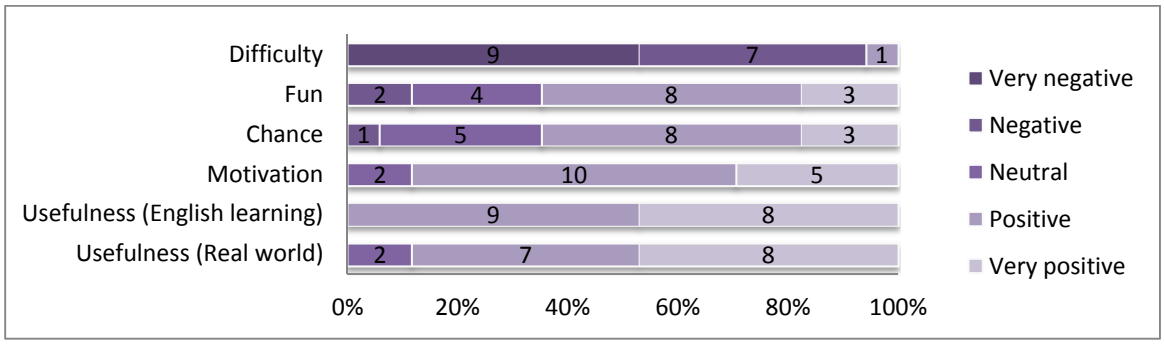



**Figure 3**. Perceptions toward TOEIC SW

Among the 16 questions, six that relate to the perceptions and evaluations to the TOEIC SW are reported in this paper. The questions sought to determine the perception of examinees on whether the test was: 1) difficult or easy; 2) fun or boring; 3) an exam they wish to re-take; 4) motivating to further study English; 5) useful in English learning; and 6) relevant after graduation. In Figure 2, the dark color toward the left signifies higher level of disagreement, whereas the light color toward the right signifies positive agreement.

Although the test was perceived as somewhat challenging ("difficulty"), the students generally regarded the test to be fun ("fun"), motivating ("motivation"), and useful ("usefulness"). Thus, they consider re-taking the test in the future ("chance"). Such results are contrary to the researcher's previous experience of students demonstrating a certain level of distress in taking the two-hour LR test. This result may be connected to the boredom, test lengthiness, and tiredness experienced by students. Particularly, the computer-based oral and writing test format was well received with a strong, positive feeling of fun during the experience.

## DISCUSSION

This study examined the viability of implementing a computer-based official test in the language education curriculum as part of career training among the science-major university students in Japan. Results confirm the appropriateness of this proposal among the students given the 500 TOEIC LR score, which likely represents the general score of university students in Japan. A certain level of gender difference that affects the adaptability to or aptitude in productive skills is likewise noted. That is, compared with males, the female students demonstrated a higher proficiency in oral skills. The expected anxiety caused by the computer-based test was not observed. Therefore, computer-based SW test is viable as an objective assessment in the language curriculum.

Further directions are hence suggested. The first point relates to the issues on computer-based test format for SW. As previously mentioned, in addition to the computer-based TOEFL that covers the four skills, which is provided by the same test agency, the IELTS is equally a widely recognized test, but incorporates a face-to-face speaking test component. Given its higher examination fees of almost USD 2500, a comparative study between TOEIC and IELTS speaking test components would be difficult. However, a study on the pros and cons of computer-based versus face-to-face test, particularly for speaking, is essential. This is particularly crucial as ICT skills and human factors are involved in the development of an objective measurement. In daily life, we are likely to speak face-to-face rather than online. However, on the global scale, such face-to-face encounters may be less evident. Even in foreign language oral assessment, we may need to assimilate both online and face-to-face competence.

The second point relates to issues of gender factors. A previous report indicated that gender produces no effect on the examinees' performance in face-to-face oral tests in IELTS (O' Loughlin, 2002). However, this may be better applied in a particular test environment, wherein examiners are screened to fit for the test objective (i.e., more gender attenuated type of examiners can be selected). A separate issue is whether or not this test situation is a more accurate reflection of the real world, specifically in communication and business contexts .

This concern could extend to other issues in Japanese universities where a generally male faculty exerts a greater influence in forming core decisions for curriculum development. A recent report collaboration by the National Institute of Science and Technology Policy and the Ministry of Education, Culture, Sports, Science, and Technology emphasized that among the academics in Engineering, female faculty ratio is as low as 3.8% in 2007 (the most recent data available), which is the lowest among disciplines in Japan. Such figure is not likely an underestimation as far as the current study is concerned. As of May 2012, the School of Engineering has only three full-time, tenured female faculty members (including myself) out of the 103 full-time faculty members (Dendai, 2013). Thus, one kind of assessment may be chosen (in this case, reading) even without a clear understanding of gender difference when assessing aptitude (in this case, speaking), which could have been indirectly disadvantaging either gender for a long time. This tendency could become more extensive as education level elevates, because additional imbalance in gender affects the ratio of the decision making body in school systems. The above issues may be examined as further research topics. English proficiency tests that have highly accumulated objective evaluations and testing could significantly contribute to this matter.

Therefore, computer-based SW test is a viable option to cope with the growing needs for objective evaluation in higher education curriculum. Further research is necessary to elucidate the underlying issues to achieve a sustainable global education.

## REFERENCES

Dendai. (2013). Dendai University Profile. Retrieved June 21, 2013 from http://web.dendai.ac.jp/about/information/.

The Institute for International Business Communication. (2012a). TOEIC proficiency scale. Retrieved April 6, 2013, from http://www.toeic.or.jp/toeic/pdf/data/proficiency.pdf

The Institute for International Business Communication. (2012b). *TOEIC Speaking/Writing data & analysis 2011*. Tokyo: The Institute for International Business Communication.

The Institute for International Business Communication. (2012c). TOEIC test data & analysis 2011. Retrieved April 6, 2013, from http://www.toeic.or.jp/toeic/pdf/data/DAA2011.pdf

Kato, M. et al. (2012). Analysis on Ratio of Women in Science in Japan. Retrieved April 7, 2013 from http://www.nistep.go.jp/achiev/ftx/jpn/mat209j/pdf/mat209j.pdf

Liao, C., Qu, Y., & Morgan, R. (2010). *The relationship of test scores measured by the TOEIC listening and reading test and TOEIC speaking and writing tests.* ( No. TC-10-13).

Miyazoe, T. & Anderson, T. (2010). Learning outcomes and students' perceptions of online writing: Simultaneous implementation of a forum, blog, and wiki in an EFL blended learning setting, System, 38(2), 185-199.

Miyazoe, T., & Anderson, T. (2012). Voice interaction online. In Q. Lin (Ed.), *Advancement in online education: Exploring the best practices* (pp. 39-67). US: Nova Science Publishing.

O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing, 19*(2), 169-192.

Obunsha. (n.d.) PassNavi: Tokyo Denki University. Retrieved April 7, 2013 from http://passnavi.evidus.com/search_univ/2850/department.html

Powers, D. E., Kim, H., Yu, F., Weng, V. Z., & VanWinkle, W. (2009). *The TOEIC speaking and writing tests: Relations to test-taker perceptions of proficiency in English.* ( No. RR-09-18, TOEIC-RR-04).ETS.

Test of English as a Foreign Language. (n.d.). TOEFL iBT. Accessible at http://www.ets.org/toefl

The International English Language Testing System. (n.d.). IELTS. Accessible at http://www.ielts.org/