



#JALT2019 • NAGOYA 11.1-11.4

## TOEIC Listening and Reading Test and Overall English Ability

Masaya Kanzaki

*Kanda University of International Studies*

### Reference Data:

Kanzaki, M. (2020). TOEIC Listening and Reading test and overall English ability. In P. Clements, A. Krause, & R. Gentry (Eds.), *Teacher efficacy, learner agency*. Tokyo: JALT.  
<https://doi.org/10.37546/JALTPCP2019-63>

The Test of English for International Communication (TOEIC) covers the four skill areas of listening, reading, speaking, and writing. This study investigates the correlations between the four skill scores of the TOEIC to determine the practicality of using the listening and reading scores to gauge overall English ability. The study was conducted over a 3-year period, during which data were collected from 260 university students. The four skill scores were significantly correlated with each other, with the highest correlation between the listening and reading scores ( $r = .73$ ), and the lowest between the listening and writing scores ( $r = .52$ ). The combined listening and reading scores were strongly correlated to the overall scores ( $r = .95$ ), which suggests that the TOEIC Listening and Reading test scores can be a good proxy for overall scores on the TOEIC.

The Test of English for International Communication (TOEIC) は、リスニング、リーディング、スピーキング、ライティングの4つの技能領域を対象にしている。本研究では、リスニングとリーディングの得点を総合的な英語力の判定に使うことの有用性を判定するため、TOEICの4技能得点間の相関を調べた。本研究は3年間に渡って行われ、260名の大学生からデータを集めた。4技能の得点間にはそれぞれ有意な相関があり、一番高い相関はリスニングとリーディングの得点間 ( $r = .73$ )、一番低い相関はリスニングとライティングの得点間 ( $r = .52$ ) であった。リスニングとリーディングの合計点は、総合点と強い相関があり、それはTOEIC Listening and Readingテストの得点がTOEICの総合点の良い代替指標になり得ることを示唆している。

Educational Testing Service (ETS) launched the Test of English for International Communication (TOEIC) in 1979 with the TOEIC Listening and Reading test (TOEIC LR) and added the TOEIC Speaking and Writing tests (TOEIC SW) in 2006 (Powers & Schmidgall, 2018). Note that the TOEIC LR is one test, with the listening and

reading sections as subcomponents, whereas the TOEIC Speaking and TOEIC Writing are two different tests, which can be taken together or separately.

The TOEIC LR and SW are both available in Japan, but there is a huge imbalance between them in terms of numbers of test-takers. In 2019, 2,205,000 test-takers in Japan took the TOEIC LR (Institute for International Business Communication, 2020a), while only 37,800 took either of the TOEIC SW tests in the same year (Institute for International Business Communication, 2020b). Because the TOEIC LR is extremely popular in Japan, these test scores are often used as a proxy for overall English ability. Some universities, for example, use the TOEIC LR scores of their students to place them into different levels of general English courses and also as indicators of the effectiveness of such courses, even when these courses put equal emphasis on the four skill areas of listening, reading, speaking, and writing. The purpose of this study is to determine whether TOEIC LR scores can serve as a proxy for overall English ability as measured by the TOEIC LR and SW.

There have been several ETS-commissioned studies with a similar aim since the 1980s. Early studies suggested that TOEIC LR scores can be a good indicator of speaking and writing ability. For example, Woodford (1982) compared TOEIC LR scores and speaking and writing test scores of Japanese English learners and found significant relationships between TOEIC Listening scores and Language Proficiency Interview scores ( $r = .83$ ,  $N = 100$ ), and between TOEIC Reading scores and direct writing exercise scores ( $r = .83$ ,  $N = 306$ ). Similarly, Wilson (1993) compared TOEIC LR scores and Language Proficiency Interview scores of 393 learners (285 from Japan, 56 from France, 42 from Mexico, and 10 from Saudi Arabia) and found a significant relationship between them ( $r = .74$ ). After the TOEIC SW tests were introduced in 2006, however, ETS publications shifted its emphasis to the notion that the abilities of speaking and writing are better assessed by the speaking and writing tests, respectively (e.g., Powers, 2010, 2013).

In addition, there have been two ETS-commissioned studies that have investigated the relationships between the four skill scores of the TOEIC. Liao, Qu, and Morgan

(2010) reported significant correlations between the four skill scores, with the highest correlation being between the listening and reading scores ( $r = .76$ ,  $N = 12,105$ ), and the lowest between the reading and speaking scores ( $r = .57$ ,  $N = 12,099$ ). Liu and Costanzo (2013) obtained similar correlations, the highest between the listening and reading scores ( $r = .726$ ,  $N = 4,935$ ), and the lowest between the listening and writing scores ( $r = .535$ ,  $N = 4,935$ ). While a correlation of over  $.70$  is usually seen to indicate an acceptable level of correlation for low-stakes tests (Hughes, 2002), both of these ETS studies maintained that the correlations in their respective studies were not sufficient to indicate a strong relationship between the different tests. In other words, they maintained that each test of the TOEIC measures distinct aspects of English language proficiency and that none of them can be used as a proxy for the others. For example, in their study, Liu and Costanzo (2013) stated that a correlation of  $.866$  or higher was necessary for one test to be able to serve as a valid predictor for another. This number was based on an index known as reduction in uncertainty, proposed by Dorans (2000, 2004); a correlation of  $.866$  is needed to reduce the uncertainty about the prediction by 50 percent. All the correlations they obtained between the four skill scores were below the threshold, and therefore they concluded that none of the four scores could be seen as a valid predictor for the others.

There have also been independent studies that have compared TOEIC LR scores with speaking and writing scores. For instance, Hirai (2008) compared TOEIC LR scores with STEP BULATS writing scores and found a strong correlation between the two ( $r = .69$ ,  $N = 559$ ). Hirai (2009) found a strong correlation between the TOEIC LR scores and STEP BULATS speaking scores ( $r = .66$ ,  $N = 493$ ). Kanzaki (2017) found a moderate correlation between TOEIC LR and TOEIC Speaking scores ( $r = .52$ ,  $N = 313$ ). Finally, Koizumi (2015) compared the four skill scores of the TOEIC ( $N = 106$ ) and reported significant correlations between them, with the highest correlation between the listening and speaking scores ( $r = .74$ ), and the lowest between the listening and writing scores ( $r = .63$ ).

The current study builds upon these previous studies by investigating the relationships between the four skill scores of the TOEIC as well as those between different sets of combined scores, with the aim of determining whether TOEIC LR scores can serve as a proxy for overall English ability as indicated by the combined TOEIC LR and SW scores.

This study was conducted to answer the following two research questions:

- RQ1. To what extent are the listening, reading, speaking, and writing scores of the TOEIC related to each other?
- RQ2. Can TOEIC LR scores serve as a proxy for overall English ability as measured by the TOEIC LR and SW?

## Method

The researcher administered the TOEIC LR and SW to the participants. The test scores were then analyzed for descriptive statistics, and Pearson's  $r$  correlations between different sets of scores were calculated.

## Participants

The participants in this study were 260 students (201 female and 59 male) who were attending a private university specializing in foreign languages in the Kanto region. They took part in the study voluntarily in exchange for a monetary reward. Informed consent was obtained from all of the participants. The objectives of the study, as well as the related procedures and requirements, were explained to the participants before they signed a consent form. The project was cleared by the university's institutional review board.

The numbers of participants in 2017, 2018, and 2019 were 98, 78, and 84, respectively. Twenty-five students participated twice, and one student participated on all three occasions. The results showed that the English proficiency levels of the repeaters had changed during the year; therefore, all data from these participants were used in the study.

Among the 260 participants, 73 were in their first year, 72 in their second, 71 in their third, and 44 in their fourth. In terms of fields of study, there were 102 English language majors, 96 international communication majors, 22 international business majors, 14 Spanish language majors, seven Portuguese language majors, seven Korean language majors, five Chinese language majors, five Indonesian language majors, and two Vietnamese language majors. As for their native languages, there were 253 Japanese speakers, four Chinese speakers, two Korean speakers, and one Vietnamese speaker. All of the participants were fluent speakers of Japanese.

Figure 1 shows how many times the participants had taken the TOEIC LR and SW prior to the study. The majority had taken the TOEIC LR, but only about 23 percent had taken the TOEIC Speaking, and even fewer (20%) had taken the TOEIC Writing.

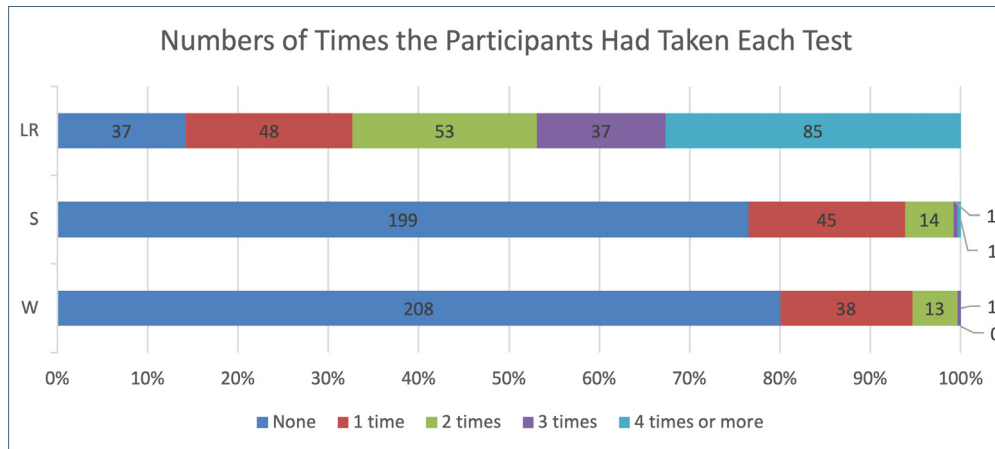


Figure 1. How many times the participants had taken the TOEIC LR (LR), TOEIC Speaking (S), and TOEIC Writing (W) prior to the study. (N = 260).

## Materials

### TOEIC Tests

The TOEIC LR is a paper-based, machine-scored examination consisting of a listening section and a reading section. Each section consists of 100 multiple-choice questions, and raw scores of between 0 and 100 are converted to scale scores of between 5 and 495. The listening section has four parts, whereas the reading section has three (see Appendix A for more details).

The TOEIC Speaking test is a computer-based examination requiring test-takers to sit in front of a computer screen while wearing a headset with a microphone. Instructions are provided on the 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 in the test, and scores are given in the range of 0 to 200 (see Appendix B for more details).

The TOEIC Writing test is also a computer-based examination. Instructions are provided on the screen, and the test-takers use a computer keyboard to type their responses, which are then sent to certified raters for evaluation. There are eight questions on the test, and scores are given in the range of 0 to 200 (see Appendix C for more details).

According to ETS publications (Educational Testing Service, 2019a; Educational Testing Service, 2019b), the reliabilities of the listening, reading, speaking, and writing scores are .90, .90, .80, and .82, respectively. According to the same sources, the standard error of measurement for the listening, reading, speaking, and writing scores are 25, 25, 13, and 17 scale points, respectively.

## Procedures

The TOEIC LR and SW were administered in the Institutional Program (IP) on campus. In 2017 and 2018, the TOEIC tests were given over 2 days; the TOEIC LR was given on the final Friday in July, and the TOEIC SW was given on the following Monday. In 2019, all the TOEIC tests were administered on September 10 despite the TOEIC LR being originally scheduled for September 9. The date was changed because most of the trains in the area were not running on September 9 due to damage caused by a typhoon the previous day. The TOEIC LR was therefore given on the same day as the TOEIC SW, which might have affected the participants' test performance.

## Results and Discussion

### Descriptive Statistics

Table 1 shows the descriptive statistics for the TOEIC scores. Note that some adjustments were made when the overall scores (the total scores) were calculated; to make the four skill scores equally weighted in the total, the SW scores were divided by 400 (the highest possible SW score) and multiplied by 990 (the highest possible LR score) before being added to the LR scores.

The average listening score was 69 points higher than the average reading score, and the average writing score was 12.43 points higher than the average speaking score.

Table 1. Descriptive Statistics for Scores (N = 260)

	Possible Scores	Mean	SD	Min	Max	Skewness	Kurtosis
L	5-495	377.46	67.70	145	495	-0.62	0.52
R	5-495	308.46	77.48	85	470	-0.29	-0.22
S	0-200	128.42	20.39	70	190	0.02	0.42
W	0-200	140.85	21.45	60	190	-0.70	0.82
LR	10-990	685.92	135.07	270	965	-0.37	-0.01
SW	0-400	269.27	37.13	140	370	-0.38	0.42
Overall	10-1980	1352.36	209.97	715.75	1855.75	-0.39	0.34

Note. L = listening scores, R = reading scores, S = speaking scores, W = writing scores, LR = listening and reading scores combined, SW = speaking and writing scores combined, Overall = LR + (SW × 990/400), SD = standard deviation, Min = lowest score, Max = highest score.

### Histograms

Figure 2 shows histograms of the four skill scores with normal distribution curves. The distributions of the listening, reading, and speaking scores are approximately normal. However, the distribution of the writing scores deviates from the normal distribution at the peak, and therefore the assumption of normality is not satisfied.

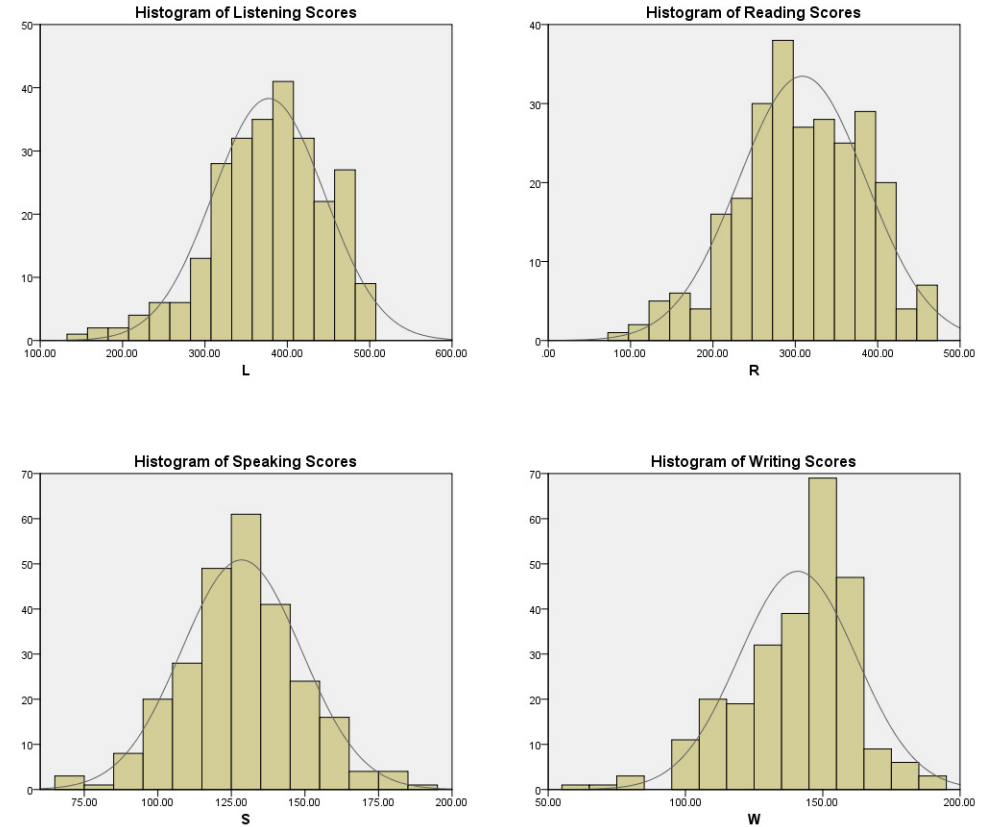


Figure 2. Histograms of the listening scores (top left), reading scores (top right), speaking scores (bottom left), and writing scores (bottom right) (N = 260).

Figure 3 shows the histograms of the LR and SW scores with normal distribution curves. The distributions of both sets of scores are reasonably close to the normal distribution, though there are some deviations.

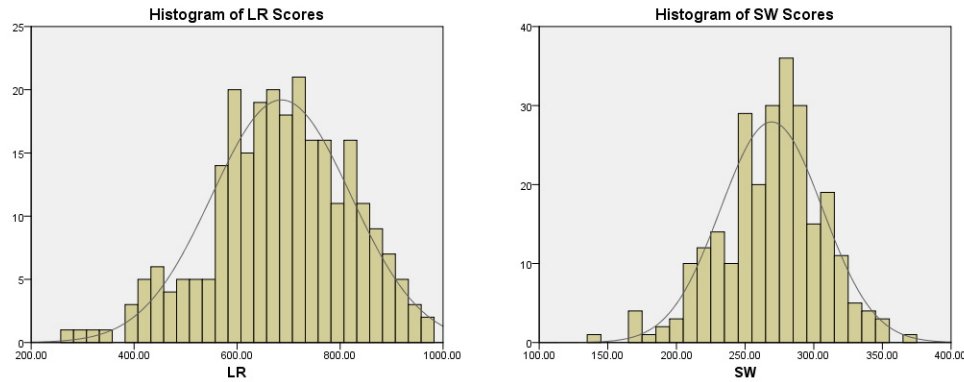


Figure 3. Histograms of the LR scores (left) and SW scores (right) ( $N = 260$ ).

Figure 4 shows the histogram of the overall scores with a normal distribution curve. The distribution of the overall scores is fairly close to the normal distribution.

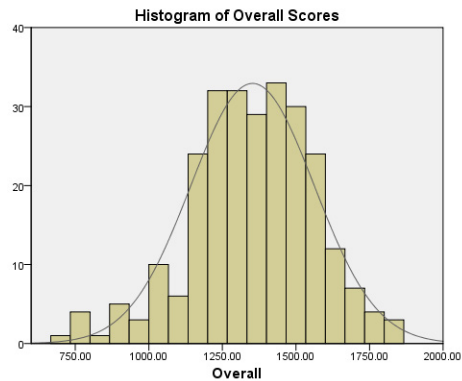


Figure 4. Histogram of the overall TOEIC scores ( $N = 260$ ).

### Correlations

As the scores did not fully meet the assumption of normality, bootstrapping was performed with 1,000 samples, and bias-corrected and accelerated bootstrap 95% confidence intervals were obtained for all correlations using IBM SPSS Bootstrapping. Table 2 shows Pearson's  $r$  correlations among the four skill scores. The scores were significantly correlated with each other ( $p < .001$ ), and the correlations were similar to those reported in Liao et al. (2010) and Liu and Costanzo (2013), though the correlations between the speaking scores and the other scores were weaker in this study than in Koizumi (2015). Furthermore, the highest correlation in the current study was between the listening and reading scores ( $r = .73$ ), while in Koizumi's study it was between the listening and speaking scores ( $r = .74$ ).

The lowest correlation in this study was between the listening and writing scores ( $r = .52$ ). In Liu and Costanzo (2013) and Koizumi (2015), the lowest correlation was also between the listening and writing scores ( $r = .535$  and  $r = .63$ , respectively), but in Liao et al. (2010), the lowest correlation was between the reading and speaking scores ( $r = .57$ ).

Table 2. Correlations Between the Four Skill Scores ( $N = 260$ )

	L	R	S	W
L	1	.73* [.66, .79]	.55* [.45, .64]	.52* [.41, .62]
R		1	.59* [.50, .67]	.64* [.56, .71]
S			1	.58* [.49, .66]
W				1

Note. L = listening scores, R = reading scores, S = speaking scores, W = writing scores. \* $p < .001$  (two-tailed). BCa bootstrap 95% CIs reported in brackets.

In response to the first research question, which asks how well the four skill scores of the TOEIC are related to each other, we can see from these results that they are correlated significantly, with Pearson's  $r$  correlations ranging from .52 to .73.

### Reduction in Uncertainty

Liu and Costanzo (2013) used a statistical index called reduction in uncertainty (RiU) as a measure to evaluate the magnitude of correlations. Dorans (2000) defined RiU as:

$$RiU = 1 - \sqrt{1 - r^2}.$$

This formula is used to calculate what proportion of uncertainty about prediction is reduced for a given correlation. For example, when  $r = 0$ , there is a zero reduction in uncertainty, and when  $r = 1$ , there is a 100 percent reduction in uncertainty. To achieve a 50 percent reduction in uncertainty, the correlation needs to be .866. Liu and Costanzo (2013) selected this number as the threshold for the acceptable level of correlation for one score to be a valid predictor for another. Following the example of Liu and Costanzo (2013), this study used RiU to evaluate the magnitude of correlations. Table 3 shows Pearson's  $r$  correlations between the TOEIC LR scores and the other scores. RiU for each correlation is given as a percentage. All Pearson's  $r$  correlations in Table 3 are significant ( $p < .001$ ).

Table 3. Correlations and Reduction in Uncertainty (RiU) Between TOEIC LR Scores and Other Scores ( $N = 260$ )

	S	W	SW	Overall
LR	.62*	.63*	.70*	.95*
	[.53, .69]	[.54, .71]	[.62, .77]	[.94, .96]
RiU	21%	22%	29%	69%

Note. S = speaking scores, W = writing scores, SW = speaking and writing scores combined, Overall = LR + (SW × 990/400), LR = listening and reading scores combined, RiU = reduction in uncertainty. \* $p < .001$  (two-tailed). BCa bootstrap 95% CIs reported in brackets.

Figure 5 shows the relationships between the LR scores and the other four sets of scores in Table 3. Notice that the points in the scatterplot of the overall and LR scores on the bottom right are tightly clustered around the regression line.

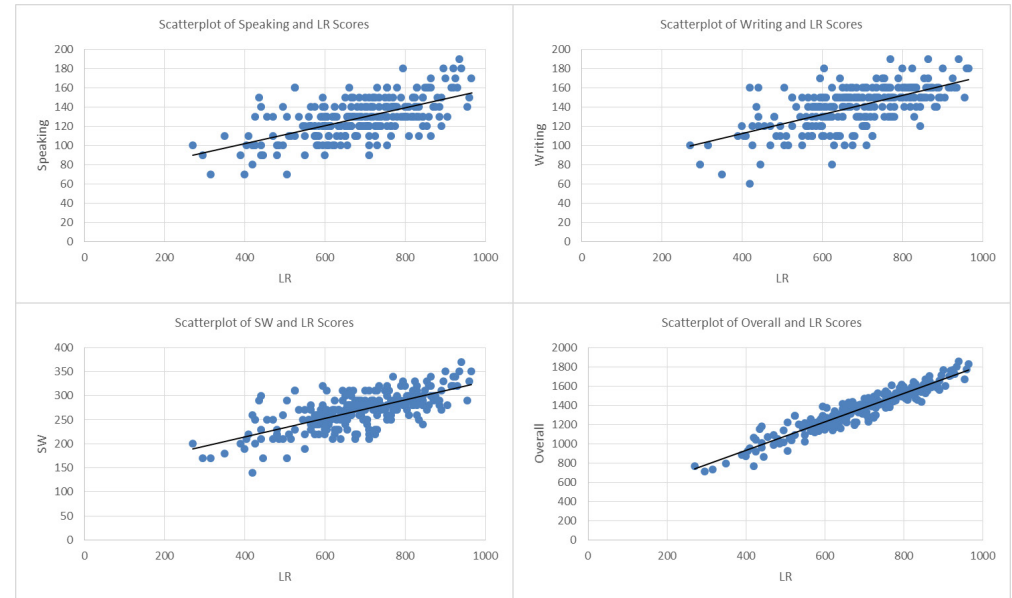


Figure 5. Scatterplots of the speaking and LR scores (top left), writing and LR scores (top right), SW and LR scores (bottom left), and overall and LR scores (bottom right) ( $N = 260$ ).

The LR scores were significantly correlated with the speaking scores ( $r = .62$ , RiU = 21%) and with the writing scores ( $r = .63$ , RiU = 22%). These correlations, however, are not high enough for predicting one score from the other with a level of certainty of 50 percent or greater.

The LR scores were more strongly correlated with the SW scores ( $r = .70$ , RiU = 29%) than with the speaking or writing scores. The correlation of .70 can be considered an acceptable level for low-stakes tests (Hughes, 2002). However, the RiU for the correlation of .70 is only 29 percent, which is far below the acceptable level of 50 percent set by Liu and Costanzo (2013). This indicates that TOEIC SW scores cannot be predicted from TOEIC LR scores with a reasonable level of certainty.

The LR scores were much more strongly correlated with the overall scores ( $r = .95$ , RiU = 69%) than with the SW scores. The correlation of .95 is 0.084 higher than the threshold of .866 set by Liu and Costanzo (2013), which suggests that TOEIC LR scores can be a

Kanzaki: TOEIC Listening and Reading Test and Overall English Ability

good predictor of overall scores. In other words, TOEIC LR scores can serve as a good proxy for overall English ability as indicated by the combined TOEIC LR and SW scores.

Powers and Schmidgall (2018) reported that some test-takers achieve high scores on the TOEIC LR despite their lack of overall ability to communicate in English, which contradicts the results of this study. One possible explanation for this is that the participants in the study were language majors and regularly had opportunities to practice speaking and writing. Consequently, they had acquired speaking and writing abilities on par with their listening and reading abilities. If learners who were less proficient at speaking and writing took the TOEIC LR and SW, the results might be different.

Another possibility is that a high overall score on the TOEIC does not always guarantee high overall English ability, as there are some aspects of overall English ability that cannot be measured by the TOEIC. For example, the TOEIC Speaking is a computer-based test, which does not examine how well a test-taker interacts with another speaker. In other words, the TOEIC Speaking does not measure interactional competence, which is an important component of the speaking construct (Galaczi, 2010; Plough, Banerjee, & Iwashita, 2018). Broadly speaking, English proficiency tests, such as the TOEFL, TOEIC, and IELTS, may not serve as good measures of test-takers' overall English ability (Brown, 2019). Therefore, it is possible that those who had high overall scores on the TOEIC might not have had high overall ability to communicate in English in real life.

With regards to the second research question, which asks whether TOEIC LR scores can serve as a proxy for overall English ability as measured by the TOEIC LR and SW, as far as the participants in this study are concerned, the TOEIC LR scores were very closely related to the combined TOEIC LR and SW scores. Therefore, TOEIC LR scores can serve as a good proxy for overall English ability as measured by the TOEIC LR and SW. One could argue that overall English ability cannot be measured accurately by the TOEIC LR and SW, but it is beyond the scope of this study to speculate on the relationship between them.

## Conclusion

From this study we can see that there is a moderate-to-strong correlation between the four skill scores of the TOEIC. This indicates that there is probably at least some overlap in the general language constructs being measured in the different skill areas of the TOEIC. However, the correlations are not high enough for one of the four skill scores to be used to predict the others with more than a 50% degree of certainty. On the other hand, because the correlation coefficient between the TOEIC LR and the combined

TOEIC LR and SW scores ( $r = .95$ ) is well over the  $r = .866$  threshold, it should be possible, at least with this group of students, to use TOEIC LR scores to predict combined TOEIC LR and SW scores. This is important because the TOEIC LR test is the one that most Japanese students take, and this could potentially allow teachers to feel more confident in using the scores for this test as a broad measure of students' overall English proficiency for the purposes of streaming and choosing level-appropriate materials for a class.

The limitation of the study is that the participants were different from typical English learners in Japan because they were more accustomed to speaking and writing in English than the average Japanese English learner. Therefore, the results may not apply to Japanese English learners in general. In future studies, it would be interesting to see what the results would be like if other types of learners took the TOEIC LR and SW.

## Supplementary Data

The TOEIC scores used in this study are available at [bit.ly/Kanzaki\\_sup](http://bit.ly/Kanzaki_sup)

## Acknowledgements

This study was supported by JSPS KAKENHI Grant Number 17K03018.

## Bio Data

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

## References

- Brown, J. D. (2019). Overall English proficiency (whatever that is). *Shiken*, 23(2), 43–47. Retrieved from [http://teval.jalt.org/sites/teval.jalt.org/files/23\\_02\\_43\\_Statistics\\_Corner.pdf](http://teval.jalt.org/sites/teval.jalt.org/files/23_02_43_Statistics_Corner.pdf)
- Dorans, N. J. (2000). *Distinctions among classes of linkages* (College Board Research Note RN-11). New York, NY: College Board. Retrieved from <https://files.eric.ed.gov/fulltext/ED562636.pdf>
- Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement*, 28(4), 227–246. <https://doi.org/10.1177/0146621604265031>
- Educational Testing Service. (2018). *Examinee handbook: Speaking & Writing tests*. Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/s/toEIC/pdf/speaking-writing-examinee-handbook.pdf>

Kanzaki: *TOEIC Listening and Reading Test and Overall English Ability*

- Educational Testing Service. (2019a). *Score user guide: TOEIC Listening and Reading test*. Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/s/toeic/pdf/toeic-listening-reading-test-user-guide.pdf>
- Educational Testing Service. (2019b). *Score user guide: TOEIC Speaking and Writing tests*. Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/s/toeic/pdf/toeic-sw-score-user-guide.pdf>
- Educational Testing Service. (2019c). *Examinee handbook: TOEIC Listening & Reading test, paper delivered*. Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/s/toeic/pdf/examinee-handbook-for-toeic-listening-reading-test-updated.pdf>
- Galaczi, E. D. (2010). Face-to-face and computer-based assessment of speaking: Challenges and opportunities. In L. Araújo (Ed.), *Computer-based assessment of foreign language speaking skills: CBA 2010* (pp. 29–51). Luxembourg: Publications Office of the European Union.
- 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. (2009). 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>
- Hughes, A. (2002). *Testing for language teachers* (2nd ed.). Cambridge, England: Cambridge University Press.
- Institute for International Business Communication. (2020a). TOEIC Listening and Reading test 受験者の推移 [The changes in the numbers of yearly test-takers of the TOEIC Listening and Reading test]. Retrieved from [https://www.iibc-global.org/library/default/toeic/official\\_data/lr/pdf/lr\\_transition\\_2019.pdf](https://www.iibc-global.org/library/default/toeic/official_data/lr/pdf/lr_transition_2019.pdf)
- Institute for International Business Communication. (2020b). TOEIC Speaking and Writing tests 受験者の推移 [The changes in the numbers of yearly test-takers of the TOEIC Speaking and Writing tests]. Retrieved from [https://www.iibc-global.org/library/default/toeic/official\\_data/sw/pdf/sw\\_transition\\_2019.pdf](https://www.iibc-global.org/library/default/toeic/official_data/sw/pdf/sw_transition_2019.pdf)
- Kanzaki, M. (2017). TOEIC Speaking test: A correlational study and test takers' reactions. In P. Clements, A. Krause, & H. Brown (Eds.), *Transformation in language education* (pp. 441-448). Tokyo: JALT. Retrieved from <http://jalt-publications.org/sites/default/files/pdf-article/jalt2016-ppc-058.pdf>
- Koizumi, R. (2015). Factor structure and four-skill profiles of the TOEIC tests among Japanese university learners of English. *ARELE: Annual Review of English Language Education in Japan*, 26, 109-124. [https://doi.org/10.20581/arele.26.0\\_109](https://doi.org/10.20581/arele.26.0_109)
- 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. In D. E. Powers (Ed.), *TOEIC compendium* (pp. 13.1-13.15). 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. In D. E. Powers (Ed.), *The research foundation for the TOEIC tests: A compendium of studies: Vol. 2* (pp. 2.1-2.25). Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/Media/Research/pdf/TC2-02.pdf>
- Plough, I., Banerjee, J., & Iwashita, N. (2018). Interactional competence: Genie out of the bottle. *Language Testing*, 35(3), 427–445. <https://doi.org/10.1177/0265532218772325>
- Powers, D. E. (2010). The case for a comprehensive, four-skills assessment of English language proficiency. In D. E. Powers (Ed.), *TOEIC compendium* (pp. 12.1-12.11). Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/Media/Research/pdf/TC-10-12.pdf>
- Powers, D. E. (2013). Assessing English-language proficiency in all four language domains: Is it really necessary? In D. E. Powers (Ed.), *The research foundation for the TOEIC tests: A compendium of studies: Vol. 2* (pp. 1.1-1.7). Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/Media/Research/pdf/TC2-01.pdf>
- Powers, D. E. & Schmidgall, J. E. (2018). The TOEIC test: A brief history. In D. E. Powers & J. E. Schmidgall (Eds.), *The research foundation for the TOEIC tests: A compendium of studies: Vol. 3* (pp. 1.1-1.5). Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/s/toeic/pdf/research-compendium.pdf>
- Wilson, K. (1993). *Relating TOEIC scores to oral proficiency interview ratings* (TOEIC Research Summaries Number 1). Princeton, NJ: Educational Testing Service. 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 Summaries). Princeton, NJ: Educational Testing Service. Retrieved from <https://www.ets.org/Media/Research/pdf/TOEIC-RS-00.pdf>



## Appendix A

### Description of Each Task in the TOEIC Listening and Reading Test

Listening Section		
Part	Task	No. of Qs
1	For each question with a photo, listen to four sentences and choose the one that best describes the image.	6
2	Listen to a question or statement followed by three responses and choose the most appropriate response.	25
3	Listen to a conversation and answer comprehension questions.	39
4	Listen to a monologue and answer comprehension questions.	30
Reading Section		
Part	Task	No. of Qs
5	Choose a word or phrase to fill in a blank in a sentence.	30
6	Choose words or phrases to fill in blanks in a passage.	16
7	Read a passage or a set of two or three passages and answer comprehension questions.	54

Note. Adapted from Educational Testing Service (2019c).

## Appendix B

### Description of Each Task in the TOEIC Speaking Test

Question No.	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.

Note. Adapted from Educational Testing Service (2018).

## Appendix C

### Description of Each Task in the TOEIC Writing Test

Question No.	Task
1-5	For each question with a photo, write a sentence based on the photo using two words or phrases given on the screen.
6, 7	Read an email message and respond to it.
8	Write an opinion essay of 300 words.

Note. Adapted from Educational Testing Service (2018).