

RESEARCH

Open Access



# Is the Common Test for University Admissions in Japan enough to measure students' general English proficiency? The case of the TOEIC Bridge

Nobuhiro Kamiya<sup>1\*</sup>

\*Correspondence:  
kamiya@fc.gpwu.ac.jp

<sup>1</sup>Department of International Communication, Gunma Prefectural Women's University, 1395-1 Kaminote, Tamamura-Machi, Sawa-Gun, Gunma 370-1193, Japan

## Abstract

This study investigated to what extent the scores of two English tests are correlated to each other, namely, the English test of the *Common Test for University Admissions* (Common Test, henceforth) in Japan and the *TOEIC Bridge*, a commercially available English test developed by *Educational Testing Service* (ETS) that measures four skills of listening, reading, speaking, and writing. Moreover, this study examined to what extent the two tests' constructs overlap from the viewpoint of L2 competence. In total, 128 university freshmen and high school seniors took the Common Test at the official venues and also the TOEIC Bridge at the researcher's university ( $n = 92$ ) or at home ( $n = 36$ ) a few months later. Results indicated that the scores of the corresponding skills are moderately correlated to each other across the two tests (Reading = .548; Listening = .646; Total = .732). Confirmatory factor analyses demonstrated that the degree of data fitting of the three models of test constructs (unitary, correlated skills, correlated tests) was statistically similar to each other. On the basis of substantive and statistical results, however, we claim that the correlated skills model should be chosen as the best-fit model and, consequently, that the productive skills should be measured in addition to the Common Test.

**Keywords:** Common Test for University Admissions, TOEIC Bridge, Language Testing, L2 competence structures

## Introduction

### Background

A majority of Japanese high school students who wish to continue studying in tertiary schools are required to take a test held nationally, organized by the *National Center for University Entrance Examinations* in Japan (National Center, henceforth). The inception of this test dates back to 1979, which was updated to the *National Center Test for University Admissions* (Center Test, henceforth) in 1990. Then, the newest version, the *Common Test for University Admissions* (Common Test, henceforth), was launched in 2021. The Common Test's English test consists of two sections of reading and listening and is

held face-to-face at around 700 testing centers all over Japan. Due to its large number of test-takers, which counts over 500,000 annually, all of the questions are in a multiple-choice format for the sake of time efficiency and fairness (see Table 1).

The *Ministry of Education, Culture, Sports, Science, and Technology* (MEXT, henceforth) originally scheduled to replace the English test within the Center Test with commercially available, private English tests of four skills when the Center Test would be switched to the Common Test due to the criticism that the Center Test measured only the receptive skills (reading and listening), despite the fact that the national curriculum guidelines stipulate that not only the receptive skills but also the productive skills (speaking and writing) must be cultivated in a balanced manner in Japanese high schools (MEXT, 2018a). This motivated Kamiya (2017) to conduct a study to investigate the score compatibility between the Center Test and one of the four-skilled private tests, namely the *TOEFL Junior Comprehensive*. The overall results supported the validity of a replacement of the former with the latter; however, the data also indicated that test-takers’ general English proficiency alone occupied the major portion of the scores, regardless of skills, across the two tests.

Due to several concerns raised during the transition period from the Center Test to the Common Test, such as difficulty in securing fairness across students with diverse economic statuses (i.e., opportunities to take private tests) and residential background (i.e., accessibility to testing venues) and score incompatibility between different private tests derived from distinct test constructs, MEXT officially announced in 2019 to postpone the replacement until 2024. However, partly because MEXT planned to use both the Common Test and private tests concurrently until then, a few major changes have been made from the Center Test to the Common Test. First, the questions to measure “pronunciation and accent” and “grammar and usage” were deleted in the reading section because, arguably, (a) the former would be better measured in a speaking test, whereas the latter in a writing test; (b) the knowledge and skills necessary to answer these questions do not emulate what is necessary for communication; and (c) they can be measured indirectly even in reading and listening sections (National Center for University Entrance Examinations, 2021). This essentially resulted in the section predominantly measuring reading comprehension skills alone. Second, in the Center Test, the scores were unevenly divided between reading (200) and listening (50) sections. But reading and listening sections each weigh evenly in the Common Test (100 & 100) (however, the actual score allocation of each of these two sections is left to the discretion of

**Table 1** Components of the English test of the Common Test

	Reading	Listening
Contents	-There is no prescribed test format common every year -Typical passages include newsletters, articles, blogs, emails, posters, and so on	1. Short sentences 2. Short conversations 3. Short monologues 4. Long monologues 5. Long conversations
Number of questions	47–49	37
Test time (Minutes)	80	30
Score range	0–100	0–100

each institution). After a series of meetings held by a special committee run by MEXT, however, MEXT announced in 2021 that the introduction of private tests at university entrance examinations would be aborted, mainly for the abovementioned reasons. Due to these changes from the Center Test to the Common Test, the applicability of the results obtained in Kamiya (2017) to the Common Test was called into question, which is the rationale for conducting the present study.

**TOEIC Bridge**

Although it would have been ideal to use the same private test as Kamiya (2017), namely, the TOEFL Junior Comprehensive, for the sake of comparability of results, at the time of conducting this study, the TOEFL Junior Comprehensive was defunct. Therefore, we needed to choose another private test. Among various candidates, we selected the TOEIC Bridge because it is designed to target beginning to lower-intermediate level learners, namely, from A1 till B1 levels of the *Common European Framework of Reference for Languages* (CEFR; Council of Europe, 2020) (Schmidgall, 2021). According to the survey conducted by MEXT, only 0.3–0.4% of third-year high school students reached B2 levels (MEXT, 2018c). Thus, we considered the test appropriate from the perspective of difficulty (but see the [Limitations](#) section).

The TOEIC Bridge originally started in 2001 as a test to measure two skills, reading and listening only. However, it had been upgraded to measure four skills starting in 2019. The test is held widely in around 35 countries (as of 2019; IIBC, personal communication, March 27, 2023). In 2021, there were 140,700 and 34,900 test-takers for listening and reading and speaking and writing, respectively (IIBC, n.d.). The test comprises four skill-based sections as shown in Table 2. The reading and listening tests can be taken either on paper or online, and all the questions are presented in a multiple-choice format. In the speaking test, test-takers record their voices through microphones in response to prompts. In the writing test, they unscramble words to complete sentences or type sentences or paragraphs. Both speaking and writing tests are held only online and their answers are evaluated by raters certified by *Educational Testing Service* (ETS), the organization that administers the TOEIC Bridge. Due to its recent introduction, to our knowledge, there have been only two attempts to compare its scores with the Common Test (and none with the Center Test). IIBC, which runs the TOEIC Bridge in Japan, has reported correlation coefficients of reading and listening scores between the TOEIC Bridge and the Common Test for 2 years in a row (IIBC, 2021, 2022). The results showed

**Table 2** Components of the TOEIC Bridge

	Listening	Reading	Speaking	Writing
Contents	1. Four pictures 2. Question-response 3. Conversations 4. Talks	1. Sentence completion 2. Text completion 3. Reading comprehension	1. Read a short text aloud 2. Describe a photograph 3. Listen and retell 4. Short interaction 5. Tell a story 6. Make and support a recommendation	1. Build a sentence 2. Write a sentence 3. Respond to a brief message 4. Write a narrative 5. Respond to an extended message
Number of questions	50	50	8	9
Test time (minutes)	25	35	15	37
Score range	30–100	30–100	30–100	30–100

moderate to strong correlations in reading ( $r=0.554, 0.592$ ), listening ( $r=0.490, 0.559$ ), and with total of both sections ( $r=0.623, 0.665$ ). However, the score of the Common Test was self-reported, not confirmed by official score reports; thus, there is a suspicion that their self-scoring might not be accurate. More importantly, their data on the TOEIC Bridge were limited in scope without speaking or writing scores. In sum, there has not been any attempt to explore the score relationships between the Common Test and four-skilled private tests.

### Models of test constructs and structure of L2 abilities

This study seeks to unveil the test constructs of the Common Test and the TOEIC Bridge through the lens of the structure of L2 abilities. Because there are a host of studies on this topic, we restricted the selection only to those that used at least one of the widely available private tests. As can be seen in Table 3, four major models of a structure of L2 abilities have been proposed as their candidates. A *unitary or unidimensional model* (Fig. 1) presupposes that L2 abilities are a single construct. Therefore, all of the test scores will be subsumed under a single, unobserved latent variable. An *uncorrelated model* (Fig. 2) presupposes that L2 competence consists of multiple, divisible, first-order variables, such as receptive and productive skills, but these variables are not correlated highly to each other. When these are highly correlated, it is called a *correlated model* (Fig. 3). Finally, when all of these first-order variables are subsumed under a single second-order variable, it is called a *high-order, second-order, or hierarchical model* (Fig. 4).

The history of these inquiries originated from Oller (1979). He analyzed multiple data sets and consistently claimed that all data could be subsumed under a single dimension. He even stated, “the current practice of many ESL programs, textbooks, and curricula of separating listening, speaking, and reading and writing activities is probably not just pointless but in fact detrimental” (p. 458). His *indivisibility hypothesis* instigated a host of ensuing explorations, most, if not all, of which criticized Oller’s use of principal component analysis (see Fouly et al., 1990), and instead, adopted more rigorous methods of confirmatory factor analysis.

Since then, among numerous kinds of private tests, a lot of attention has been paid to the TOEFL, probably due to its large number of test-takers with various backgrounds (around a million test-takers in around 160 countries annually; ETS Japan, personal communication, May 30, 2023). This convenient feature makes it easy to secure strong statistical power and to compare data among multiple diverse groups (Stricker & Rock, 2008). Through this line of research, a clear trend has appeared, which is that either a correlated or a higher-order model is acceptable, rejecting unitary and uncorrelated models (except Wilson, 2000, but see In’nami & Koizumi, 2012, for its possible reasons). When the number of first-order factors is two, a higher-order model cannot be identified. When it is three, these two models are statistically indistinguishable from each other; in such a case, a higher-order model is chosen based on the principle of parsimony. Thus, it is often impractical to decide which one is the best fit. As the consensus has been almost reached on the structure of L2 abilities, the momentum toward identifying the structure of L2 abilities has waned, and we witness a shift of studies toward validating newly made tests, such as the TOEFL iBT and the TEAP when they came out in public. Pertinent to the present study, however, is that there has been only one study

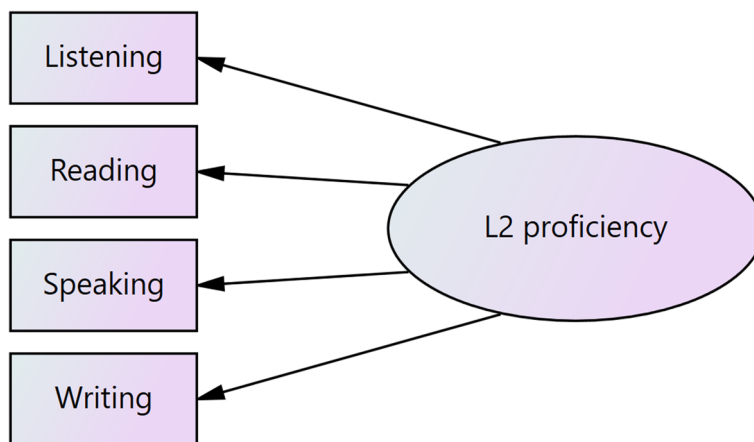
**Table 3** List of studies on test constructs and structure of L2 abilities using private tests

Study	Participants	Participants' L1	Private Tests	Unitary	Uncorrelated	Correlated	Higher-order
Bachman et al. (1995)	Various	Various	Cambridge Examinations & TOEFL PBT	×	×	△	○
Fouly et al. (1990)	University students	Various	TOEFL PBT + α	×	×	○	○
Gu (2015)	Aged 11–15	Various	TOEFL JC	×	×	×	○
Hale et al. (1988)	Various	Various	TOEFL PBT	×	×	○	-
Hale et al. (1989)	Various	Various	TOEFL PBT	×	×	○	-
In'nami and Koizumi (2012)	University students	Mostly Japanese	TOEIC LR	×	×	○	-
In'nami et al. (2016)	University students	Japanese	TEAP & TOEFL iBT	×	×	×	○
Kamiya (2017)	High school students	Japanese	TOEFL JC & Center Test	△	×	○	-
Kunnan (1995)	Various	Various	TOEFL PBT	×	×	○	×
Manning (1987)	Various	Various	TOEFL PBT	×	×	○	-
Nakamura (2022)	University students	Japanese	TEAP	×	×	×	○
Oller (1979)	Various	Various	TOEFL PBT	○	×	×	×
Oller and Hinofotis (1980)	University students	Various	TOEFL PBT	○	×	×	×
Stricker and Rock (2008)	Various	Various	TOEFL iBT	×	×	△	○
Sawaki et al. (2008)	Various	Various	TOEFL iBT	×	×	△	○
Sawaki et al. (2009)	Various	Various	TOEFL iBT	×	×	△	○
Sawaki and Sinharay (2018)	Various	Various	TOEFL iBT	×	-	○	△
Shin (2005)	Various	Various	TOEFL PBT	×	×	△	○
Swinton and Powers (1980)	Various	Various	TOEFL PBT	×	×	○	-
Wilson (2000)	Various	Japanese & Korean	TOEIC LR	×	○	×	×

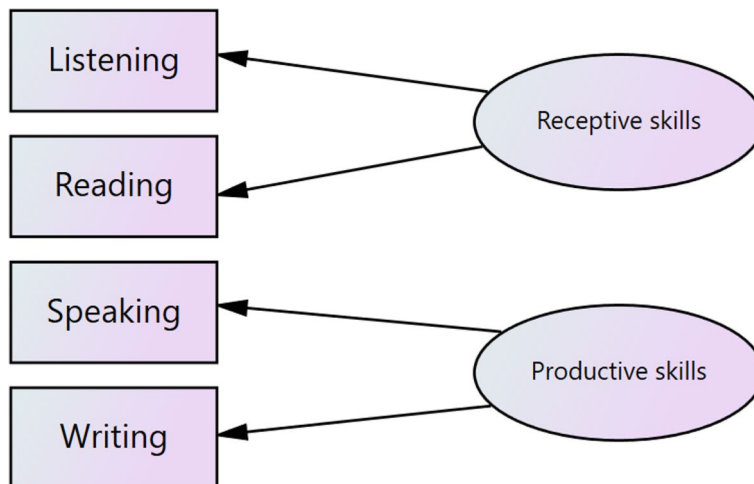
The table is not intended to be an exhaustive list of previous studies on this subject. *TEAP* Test of English for Academic Purposes (available only for Japanese university applicants), *TOEFL JC* TOEFL Junior Comprehensive, *TOEIC LR* TOEIC Listening & Reading. "△" denotes that this model was chosen as the second best. "-" denotes that this model was not considered. In the case of higher-order model, it was not considered due to the fact that the number of first-order factors was less than four (see the main text for its rationales)

on the TOEIC (In'nami & Koizumi, 2012), but only for listening and reading, not for speaking or writing, and none for the TOEIC Bridge.

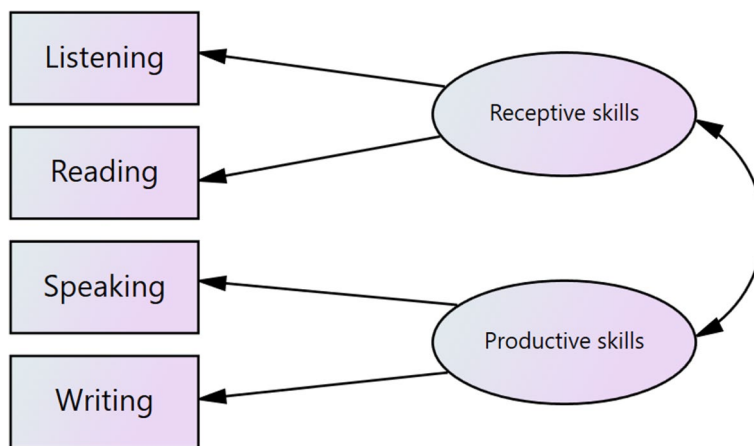
Somewhat unpredictably, in Kamiya (2017), which targeted the Center Test and the TOEFL Junior Comprehensive, although the correlated model was chosen as the best-fit model, the unitary model was found to be almost equally a good fit as well. However, although not reported in the article, when only the data of the TOEFL Junior Comprehensive were extracted and analyzed, the correlated model with two variables of receptive and productive skills was shown to be clearly better than the unitary model (e.g., SRMR = 0.0040 and 0.0132, respectively). This implies that the Center Test, especially its reading section, was measuring general English proficiency rather than reading abilities alone, which skewed the whole data set to close on the unitary model. If so, since the pronunciation and grammar sections in the Center Test were excluded at the transition



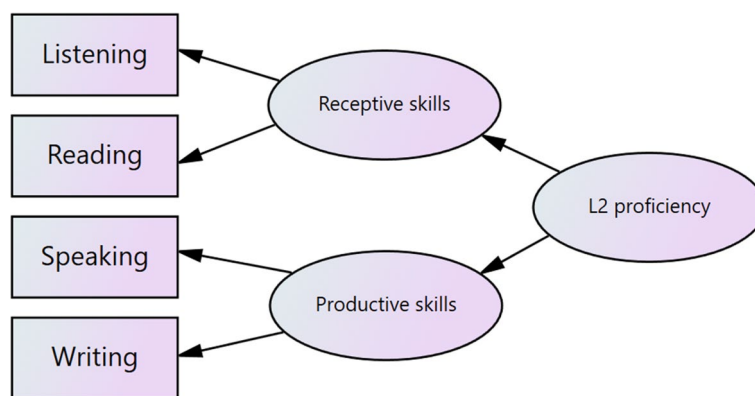
**Fig. 1** Sample of unitary or unidimensional model



**Fig. 2** Sample of uncorrelated model



**Fig. 3** Sample of correlated model



**Fig. 4** Sample of high-order, second-order, or hierarchical model

to the Common Test, we can predict that, in the case of the Common Test, the correlated model will demonstrate a better fit when compared to the unitary model.

**Research questions**

Some additional explanations are necessary regarding the models that we consider. First, owing to the fact that there are only two first-order variables (e.g., receptive and productive skills), this study is incapable of identifying a higher-order model. Second, uncorrelated models will not be considered as they were unidentified in Kamiya (2017). Third, there will be two versions of correlated models to be examined, namely, skill-based and test-based models. The former is following the convention of past literature (receptive and productive skills). For the latter, whereas the Common Test must strictly follow the national curriculum guidelines (MEXT, 2018a), the TOEIC Bridge has no such restriction for its worldwide administration. Therefore, dividing the test scores into test types, rather than skills, may produce a better fit. Therefore, the test-based model was additionally considered. In sum, the present study was guided by the following two research questions.

1. How are the scores of the Common Test and the TOEIC Bridge correlated with each other?
2. Which of the three models (unitary, correlated skills, and correlated tests) best represents the test constructs of the Common Test and the TOEIC Bridge?

**Methods**

**Participants**

The original plan was to recruit only high school students, following the procedure in Kamiya (2017); however, due to the COVID-19 pandemic, the researcher could not get permission to do so from the university. Therefore, the students of the researcher’s university were invited to participate throughout the study period. In the final year, though, this restriction was lifted, so a group of high school students participated. In total, 128 Japanese learners of English aged 18 or 19 participated in this study (118 females, 10 males), which consisted of four groups as shown in Table 4.

The university students were all freshmen. Efforts were made to recruit students with diverse majors in order to secure a wide range of English proficiencies for ensuring high reliability of analysis (Mizumoto, 2014): international communication ( $n = 36$ ), Japanese literature ( $n = 19$ ), English literature ( $n = 18$ ), fine arts ( $n = 10$ ), and liberal arts ( $n = 10$ ). The high school students were from five schools in the same district. This study was approved by the ethics committee of the researcher’s university and all of the participants agreed to participate in the study and read and signed the consent form.

**Procedure and analyses**

For university students, solicitation emails were sent to all freshmen right after enrollment from the researcher. For high school students, the researcher first contacted the principals of 10 high schools for permission for recruitment of their third-year students. After all of them agreed, the flier was distributed to them either face-to-face (paper) or online (PDF) from a teacher in each school. For both university and high school students, those students who were interested in participation voluntarily contacted the researcher.

All of the participants took the Common Test at the official venues in January. High school students took the TOEIC Bridge in the following March, in the same month of graduation. University students took it in the following May, a month after enrollment. Three groups took the TOEIC Bridge in a CALL Lab at the researcher’s university. Due to the spread of the pandemic, following the regulations imposed by the university, those in 2021 needed to take it online at home using Zoom with a camera on, invigilated by the researcher. The scores of the Common Test were confirmed by the official score reports provided by the National Center. The scores of the TOEIC Bridge were confirmed by the official score reports provided by IIBC.

Analyses were conducted in three steps. First, Pearson’s correlations were conducted in order to see to what extent the score of each section and total of the Common Test and the TOEIC Bridge are correlated. Second, confirmatory factor analyses (with maximum likelihood estimation) were conducted to detect the model that best fits the current data. Finally, chi-square difference tests were conducted in order to compare model fits. In accord with the recommendation not to conduct an exploratory factor analysis prior to a confirmatory factor analysis on the same data set, as it leads to model overfit (overly optimistic modeling) (e.g., Fokkema & Greiff, 2017), it was not implemented.

**Table 4** Demographic information of participants

Year	Participants	N	Male	Female	Administration dates	
					Common Test	TOEIC Bridge
2021	University students	36	0	36	January 16	May 29, 30 (online)
2022	University students	32	0	32	January 15	May 7
2023	University students	25	0	25	January 14	May 13, 16, 17
2023	High school students	35	10	25	January 14	March 14, 16



## Results

### Descriptive statistics

Although the National Center tries to equalize the test difficulty of the Common Test across multiple years, there are variations in the mean scores because the scores are not adjusted to secure the same level of difficulty, a practice seen in the TOEFL and the TOEIC, called *score equating* (Livingston, 2014). Therefore, the mean scores of the Common Test need to be examined to check whether the three versions of the Common Test used in the present study had approximately the same difficulty level. Table 5 shows the means and standard deviations of the Common Test across Japan in the 3 years when this study was conducted (National Center for University Entrance Examinations, n.d.-d). Although the mean scores admittedly varied among these 3 years, no score adjustment was made and all the scores were aggregated for the following reasons. First, the National Center stipulates that scores will be adjusted when averages across subjects differ by over 20 (although this applies only to those tests conducted in the same year, and English is not a subject for the adjustment), but the widest gap in the present data was around eight ( $61.80 - 53.81 = 7.99$  for reading), which is much lower than the benchmark of 20. Second, one-way ANOVAs confirmed that the total scores of the Common Test ( $p = 0.100$ ) and the TOEIC Bridge ( $p = 0.113$ ) were not significantly different across the four groups of participants. Because the score equivalency of the TOEIC Bridge is established by score equating (Livingston, 2014), the English proficiency of these four groups can be assumed to be homogeneous. Since the groups' scores on the Common Test were also similar, the Common Test's score equivalency was presumed. Table 6 shows the descriptive statistics of the scores of the participants in this study.

### Pearson's product-moment correlation matrix

Table 7 shows the results of Pearson's product-moment correlation matrix. As expected, the highest correlations are observed between the scores of each section and its total score within the same test, except for the TOEIC Bridge writing, which marked a rather low coefficient ( $r = 0.661$ ). More importantly, the scores of the corresponding skills also demonstrated relatively high coefficients across the two tests (reading = 0.548; listening = 0.646). Moreover, the total score showed an even higher coefficient ( $r = 0.732$ ) (see Figs. 5, 6, and 7).

**Table 5** Descriptive statistics of the Common Tests across Japan

	Reading	Listening	Total
Full score	100	100	200
2021	58.80 (21.44)	56.16 (16.45)	114.96
2022	61.80 (20.30)	59.45 (18.16)	121.25
2023	53.81 (20.99)	62.35 (18.82)	116.16
Mean (standard deviation)			

**Table 6** Descriptive statistics of scores

	Common Test			TOEIC Bridge				
	Reading	Listening	Total	Listening	Reading	Speaking	Writing	Total
Full score	100	100	200	50	50	50	50	200
Mean	70.05	74.13	144.17	39.30	43.70	40.05	45.69	168.73
SD	14.76	11.63	22.99	6.77	5.69	7.63	4.40	19.75
Max	100	100	200	50	50	50	50	199
Min	28	43	75	20	22	15	27	102
Skewness	−.39	−.46	−.39	−.59	−1.48	−1.13	−1.62	−.99
Kurtosis	−.11	−.04	.23	−.15	2.64	1.11	2.78	.91

SD Standard deviation

**Table 7** Results of Pearson’s product-moment correlation matrix

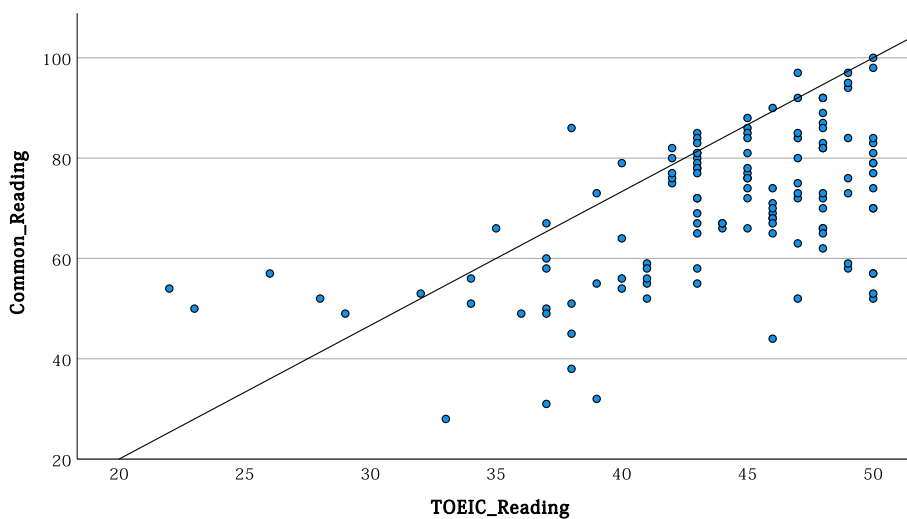
	Common Test		TOEIC Bridge				
	Listening	Total	Listening	Reading	Speaking	Writing	Total
Common Test							
Reading	.512	.901	.492	.548	.540	.350	.613
Listening	-	.834	.646	.573	.514	.374	.668
Total		-	.643	.642	.606	.414	.732
TOEIC Bridge							
Listening			-	.687	.582	.395	.854
Reading				-	.519	.462	.827
Speaking					-	.440	.833
Writing						-	.661

All correlations are significant at  $p < .001$

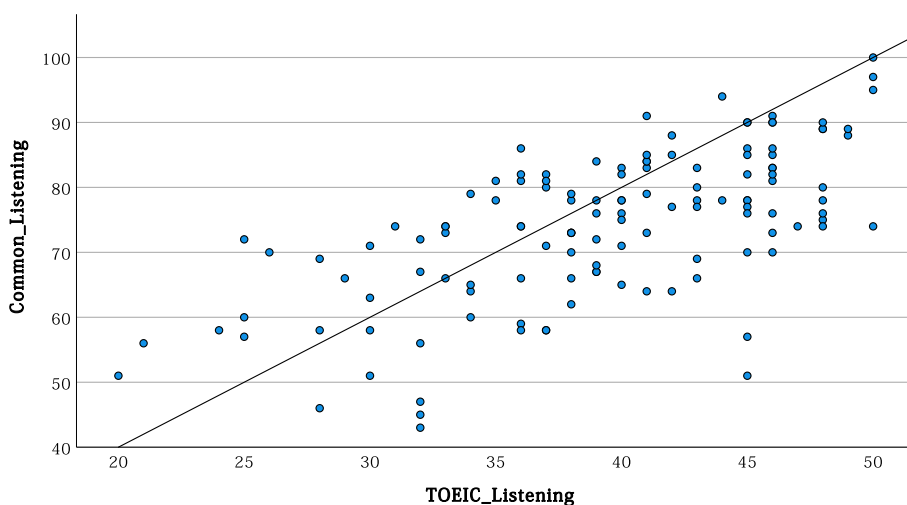
**Confirmatory factor analyses**

The results of model fits are presented in Table 8. The results of standardized regression weights are graphically presented in Figs. 8, 9, and 10. The results of unstandardized regression weights, variances, and squared multiple correlations can be obtained in the Additional file 1.

According to Table 8, all of the three models seem to be satisfactorily well fit since most of the indices meet the minimum requirement levels (i.e., RMSEA < 0.08, SRMR < 0.05, CFI > 0.95, NFI > 0.95, TLI > 0.95; Byrne, 2016). Because the unitary model is nested in the other two models, chi-square difference tests were performed. The results demonstrate that the unitary model was not a significantly better fit than the correlated skills model ( $p = 0.182$ ) nor the correlated tests model ( $p = 0.688$ ). That is to say, statistically speaking, all of these three models are equally well fit.



**Fig. 5** Scatter plot of reading scores ( $r = .548$ )

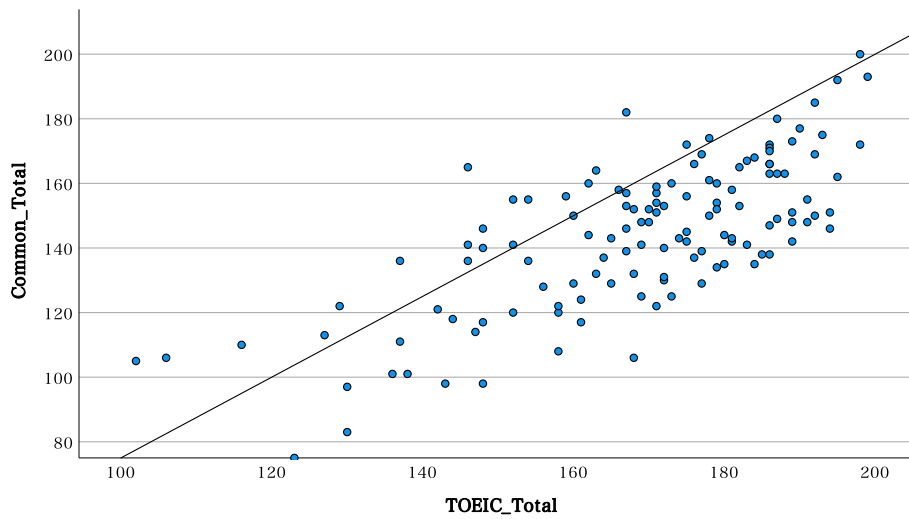


**Fig. 6** Scatter plot of listening scores ( $r = .646$ )

## Discussion

### RQ1: Correlations between the Common Test and the TOEIC Bridge

The results of Pearson’s product-moment correlations showed that the scores of the Common Test and the TOEIC Bridge are moderately correlated for reading ( $r = 0.548$ ), listening ( $r = 0.646$ ), and also total ( $r = .732$ ). A higher correlation coefficient for the total score than for the individual skill (reading or listening) is commonly observed in other studies in which the total score is derived from the sum of the two skills (reading and listening) (IIBC, 2021, 2022) and of the four skills (Kamiya, 2017). Inspecting Figs. 5, 6, and 7, we surmise that this is probably because the total score better reflects the participants’ English proficiency owing to the reduced amount of measurement error by combining the scores of multiple tests, which decreases the deviations from the regression line. Plonsky and Oswald (2014) proposed new benchmarks of correlation coefficients

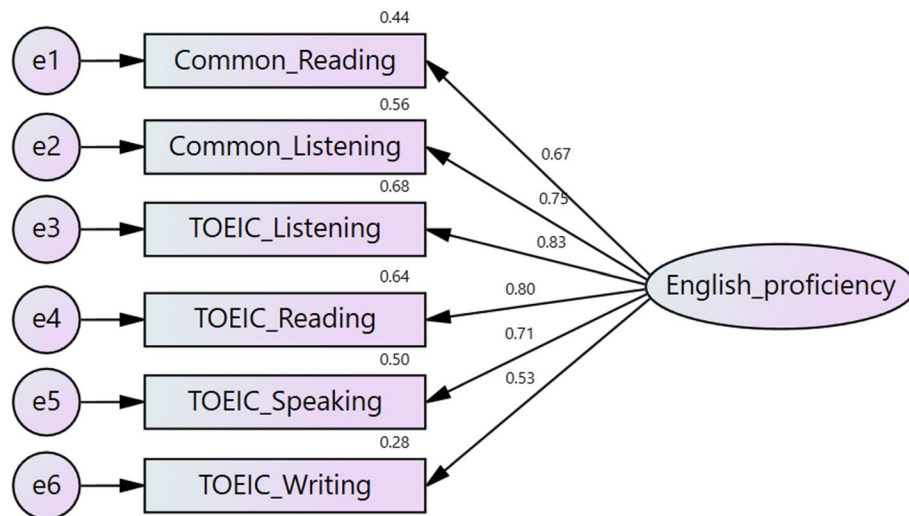


**Fig. 7** Scatter plot of total scores ( $r = .732$ )

**Table 8** Results of model fits

	$\chi^2$	df	RMSEA [CI]	SRMR	CFI	NFI	TLI
Unitary	13.252	9	.061 [0, .126]	.032	.986	.960	.977
Skills	11.476	8	.058 [0, .128]	.029	.989	.965	.979
Tests	13.091	8	.071 [0, .137]	.032	.984	.960	.969

df Degree of freedom, CI 90% confidence interval



**Fig. 8** Standardized regression weights of the unitary model

for L2 studies, with 0.25 being weak, 0.40 being medium, and 0.60 being large. According to these criteria, all of these correlations can be said to be medium to large. These figures are roughly equal to those correlation coefficients obtained for the Common

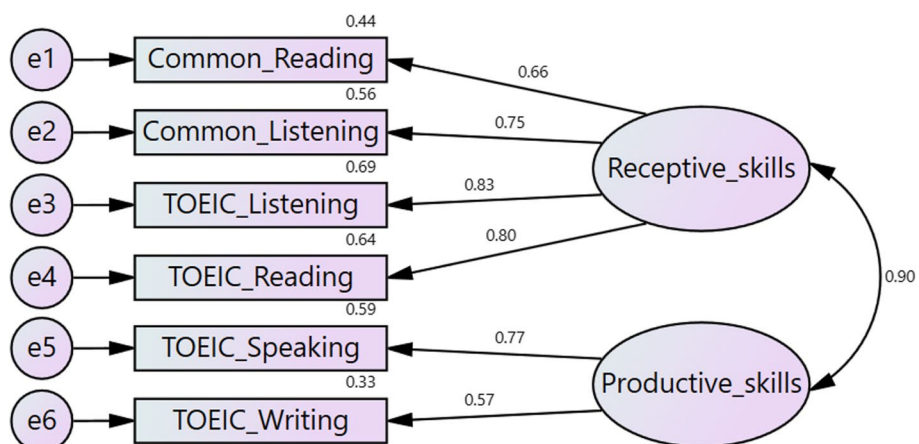


Fig. 9 Standardized regression weights of correlated skills model

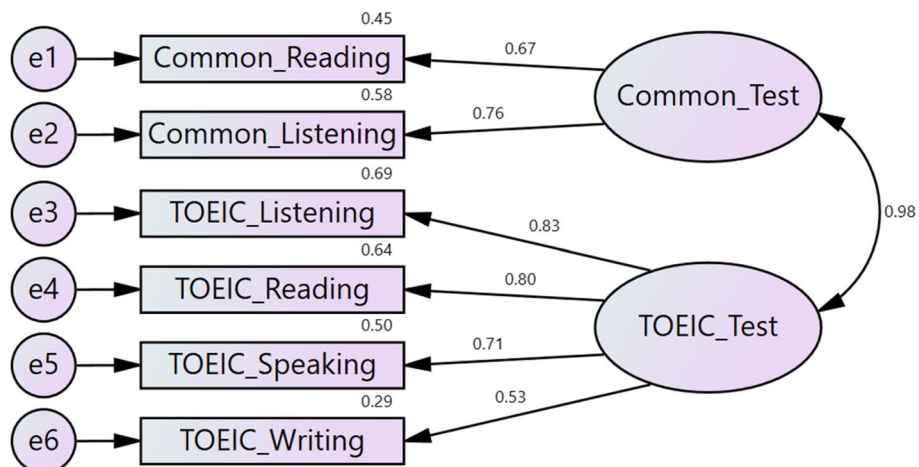


Fig. 10 Standardized regression weights of correlated tests model

Table 9 Correlation coefficients between Center Test or Common Test and private tests

Study	Entrance exam	Private test	Reading	Listening	Total <sup>a</sup>
Otsu (2014)	Center Test	TOEIC	-	-	.80
		TOEFL ITP	-	-	.70
Eiken Foundation of Japan (2015)	Center Test	TEAP	-	-	.80
		STEP Eiken <sup>b</sup>	-	-	.89
Kamiya (2017)	Center Test	TOEFL Junior Comprehensive	.78	.75	.86
IIBC (2021)	Common Test	TOEIC Bridge	.55	.49	.62
IIBC (2022)	Common Test	TOEIC Bridge	.59	.56	.67
Present study	Common Test	TOEIC Bridge	.55	.65	.73

<sup>a</sup> Total refers to the added score of reading and listening except for Kamiya (2017) and the present study, which refers to the added score of the four skills

<sup>b</sup> STEP Eiken is a test to measure English proficiency conducted in Japan. The test is currently divided into seven levels, five of which measure four skills whereas the other two (the lowest levels) measure the receptive skills only

Test or the Center Test against several private tests, as can be seen in Table 9. However, according to Dorans (2004), the correlation coefficient of 0.866 is minimally necessary for a test to be replaced by another. Although the interchangeability of the Common Test and the TOEIC Bridge is not the objective of our inquiry, at least from a psychometric standpoint, their test constructs are distinct enough to warrant further examination of their reasons. For the sake of comparisons, we summarized the selected specifications of these two tests in Table 10, taken from multiple sources for the Common Test (National Center for University Entrance Examinations, n.d.-c, n.d.-d) and the TOEIC Bridge (Everson et al., 2021; Schmidgall, 2021; Schmidgall et al., 2019, 2021). From this table, it is clear that although these tests share some commonalities, such as the objective to deal with communication in real daily life contexts and CEFR levels to be measured (A1-B1), there are a number of differences between them.

First, the Common Test presumably targets the life of high school students because the questions are made considering the “situations in which students learn in the classroom, discover problems in their social and daily lives” (National Center for University Entrance Examinations, n.d.-b, pp. 1–2). In contrast, the TOEIC Bridge deals with *adult* life, which makes some of the questions irrelevant to high school students. For instance, IIBC provides sample questions about the TOEIC Bridge on their website (IIBC, n.d.). A listening question plays, “What color is your car?” A reading question (fill in the blank) reads, “We have received your order for twelve yellow roses.” A speaking question asks to summarize the announcement made by a company president at a staff meeting. A writing question asks to reply to a question, “what types of training or education do you think people will need to get well-paid jobs in the future?” High school students would probably never encounter a situation to be exposed or use any of these sentences in their daily lives even in the L1 (Japanese).

Second, the Common Test is expected to follow the national curriculum guidelines (MEXT, 2018a) whereas the TOEIC Bridge bears no such obligation. Thus, in the latter, some of the linguistic items may go beyond what is supposed to be covered in the former. For example, a sample reading section of the TOEIC Bridge (Educational Testing Service, 2020) has a choice of “An employee’s retirement” in a multiple-choice question. Although this is a high-frequency phrase in the workplace, high school students may not be familiar with such a phrase.

Finally, the Common Test mainly consists of American English, and to a much lesser extent, British English and Japanese-accented English (National Center for University

**Table 10** Selected specifications of the Common Test and the TOEIC Bridge

	Common Test	TOEIC Bridge
Objective	To measure the skills to use knowledge of English vocabulary, expressions, grammar, and language functions appropriately in real-life communication according to the purpose, situation, and circumstance	To measure English language proficiency in the context of everyday adult life
Dialects	The USA, the UK, Japan	The USA, the UK, Canada, Australia
Restrictions of guidelines	Yes	No
CEFR	A1-B1	A1-B1

Entrance Examinations, n.d.-a). This is because Japanese students are used to American English for two reasons: (a) most, if not all, English textbooks used in Japanese schools are written in American English (e.g., Mitsumura Toshio, n.d.; Tokyo Shoseki, 2018), and (b) the majority of teachers who come from abroad to teach English are Americans. For instance, in one of the largest programs for hiring foreign teachers, the JET (Japan Exchange and Teaching) Programme, as of 2023, Americans comprise 55% of the entire faculty (Council of Local Authorities for International Relations, 2015). In regard to the use of Japanese-accented English, because sharing the same L1 between speakers and listeners is known to facilitate comprehension (e.g., Tergujeff, 2023), high school students should find Japanese-accented English easier to comprehend. On the other hand, in addition to American and British English, the TOEIC Bridge contains those dialects of Canada and Australia, both of which are lacking in the Common Test. Because high school students are unfamiliar with Canadian and Australian English, they may have difficulty understanding them compared to American English.

All in all, these discrepancies in specifications between these two tests may have yielded correlation coefficients not high enough to be replaceable. Looking at Table 9 we find that most of the correlation coefficients in the previous and the present studies did not reach the benchmark of 0.866 (Dorans, 2004). This makes sense because the test specifications required for university matriculations imposed on Japanese high school students should be quite distinct from those for assessing the English proficiencies of test-takers all over the world.

#### **RQ2: Test constructs of the Common Test and the TOEIC Bridge**

The results of confirmatory factor analyses revealed that the three models compared (unitary, correlated skills, correlated tests) explained the data equally well (see Table 8). The correlated skills model and the correlated tests model have the same degree of freedom; therefore, a chi-square difference test cannot be conducted between these two models. Comparing the values in each index, the correlated skills model seems to be superior to the correlated tests model (e.g., RMSEA = 0.058 and 0.071, respectively). Moreover, in the correlated tests model, the two latent variables (Common\_Test and TOEIC\_Test) are highly correlated to each other with the correlation coefficient being 0.98. Traditionally, when the correlation coefficient is over 0.9, they are regarded as being statistically indistinguishable (Gu, 2015). Therefore, we deem that the correlated tests model should be rejected.

Following this notion, among the two models left (unitary and correlated skills), if we follow the principle of parsimony, the simpler model with more degrees of freedom (unitary) should be chosen over a more saturated model with fewer degrees of freedom (correlated skills). Yet, because this homogenous result may be ascribed to the small sample size (see the [Limitations](#) section), we must consider other criteria for judgments, and we deem that the correlated skills model is the best-fit model for the following reasons.

First, as can be seen in Table 3, Oller's studies are the only ones that fully supported the unitary model, and a large body of literature following them uniformly negates it. Based on the accumulation of such empirical evidence, it is more natural to disregard its viability.

Second, since the correlation coefficient of the two latent variables (Receptive\_skills and Productive\_skills) is not over 0.9, due to the abovementioned reason, it is more valid to set up these two constructs separately, rather than combining them into a single construct.

Finally, in order to find out to what extent the TOEIC Bridge was capable of distinguishing the receptive and productive skills, we performed confirmatory factor analyses of two models (unitary and correlated skills) only with the data of the TOEIC Bridge. Table 11 shows the results of model fits. To our surprise, unlike the TOEFL Junior Comprehensive, the TOEIC Bridge does not seem to well distinguish between receptive and productive skills. A chi-square difference test also confirmed that these two models are statistically homogenous ( $p=0.128$ ). This may be because of the nature of some of the questions for speaking and writing tests, in which integrative skills are necessary to answer such questions (see the [Limitations](#) section for another possible reason). Looking at Table 11, the unitary model is superior to the correlated skills models at two indices (RMSEA and TLI) whereas the opposite is true for the other three (SRMR, CFI, and NFI). Therefore, there is no consistent pattern as to the superiority of either model. Compared to that, when combined with the scores of the Common Test, all the indices consistently favor the correlated skills model because the correlated skills model recorded (a) the lowest values in RMSEA and SRMR and (b) the highest values in CFI, NFI, and TLI (Table 8). Adding the scores of the Common Test strengthened the model fit of the correlated skills model indicating that the Common Test itself measured receptive skills rather than general English proficiency.

**Limitations**

There were several limitations in the present study. First, there was a time gap of two (high school students) to four (university students) months between the Common Test and the TOEIC Bridge. This was unavoidable due to logistic reasons. It is unknown whether the participants’ English proficiency changed, and if so, how much, during this period.

Second, although we aimed to recruit roughly the same number of participants as Kamiya (2017) ( $n = 144$ ), due to the limitation on budget, this was unfeasible. This could be a reason why all three models were found to be homogeneously well fit to the data. Thus, the results should be interpreted cautiously.

Third, although the level of the TOEIC Bridge is appropriate for the majority of high school students on the national level, it may have been too easy for the participants in this study. Among the 93 university students, 54 (58.1%) were English majors. All the high school students were recruited from high-level schools in the district. According to Table 6, the accuracy rate reached approximately 80% across all the skills and all of the

**Table 11** Results of model fits of the TOEIC Bridge

	$\chi^2$	df	RMSEA [CI]	SRMR	CFI	NFI	TLI
Unitary	5.325	2	.114 [0, .237]	.0319	.981	.970	.942
Skills	3.009	1	.126 [0, .300]	.0210	.988	.983	.930

df Degree of freedom, CI 90% confidence interval



data were negatively skewed, which may have weakened its discriminative power. Moreover, the mean score of the writing section was as high as 45.7/50, which may explain why this section had the lowest factor loading in all the models.

Fourth, 36 participants who participated in the first year of this project needed to take the TOEIC Bridge online due to the pandemic. Independent *t*-tests indicate that they scored higher than those who took it face-to-face for speaking ( $p=0.010$ ) and total scores ( $p=0.049$ ). Although speculatively, the online participants may have felt less anxious when speaking English aloud at home alone compared to those who took it in a computer lab with the presence of other students around them hearing their voices from each other. Although this decision was beyond the researcher’s control, there is some doubt about their score equivalency. The pandemic facilitated online administration of even high-stakes tests, such as the TOEFL iBT, but to my knowledge, there has not been any systematic attempt to validate the score comparability of the speaking section between online and face-to-face. This point may be worthy of being addressed in future studies.

**Conclusion**

Commotion ensued around 10 years ago when the idea of replacing the Center Test by private tests of four skills was introduced by MEXT. After countless heated arguments, this idea has been officially aborted. The introduction of assessment on the productive skills, be they speaking and/or writing, to university entrance examinations, does not seem to be happening anytime soon.

However, the picture does not seem completely bleak for three reasons. First, a speaking test, called ESAT-J (English Speaking Achievement Test for Junior High School Students) has been introduced into an entrance examination for all public senior high schools in Tokyo from November 2022 despite the fact that speaking skills are “the most logistically challenging and controversial to assess” (O’Sullivan et al., 2022, p. 12) among the four skills. Second, a few universities are devising their own ways to measure students’ speaking skills at their entrance examinations (Committee for Selection of Good Practices in University Admissions, 2022). Third, a growing number of universities are now utilizing the scores of private tests of four skills for entrance examinations by, for instance, requiring a certain score for applying or adding extra points to the score of the test conducted by each university (Kawaijuku Education Institution, n.d.). Thus, albeit slowly, we are moving forward from the measurements of receptive skills only toward that of four skills in line with the high school English curriculum guidelines (MEXT, 2018b), which stipulates that all of the four skills must be cultivated in a balanced manner. We truly hope that the National Center will create a reliable and valid measurement of four skills someday. But until that day comes, a viable solution seems to have each university implement an assessment of productive skills on their own in addition to the Common Test.

**Abbreviations**

CEFR	Common European Framework of Reference for Languages
Common Test	Common Test for University Admissions
ETS	Educational Testing Service
MEXT	Ministry of Education, Culture, Sports, Science, and Technology
National Center	National Center for University Entrance Examinations in Japan
Center Test	National Center Test for University Admissions

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40468-024-00272-6>.

**Additional file 1.** Results of unstandardized regression weights, variances, and squared multiple correlations.

### Acknowledgements

First, this study would not be feasible without the continuous support from Ayaka Shibata, a staffer of IIBC, which runs the TOEIC Bridge in Japan. Second, I would like to express my deepest gratitude to Paula Winke, Rie Koizumi, and Yo In'nami who gave me valuable comments on an earlier version of this paper. Third, I am grateful to Mark Freiermuth and Neal Snape for proofreading this manuscript. Lastly, I thank all of the participants and the high school principals who kindly allowed me to recruit students at their schools and, not to mention, the editor and anonymous reviewers for their constructive feedback.

### Author's contributions

NK singly designed the study, collected the data, performed the statistical analysis, drafted the manuscript, and approved the final manuscript.

### Funding

This work was supported by JSPS KAKENHI Grant Number JP21K00764.

### Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

## Declarations

### Competing interests

The author declares that he has no competing interests.

Received: 9 October 2023 Accepted: 5 January 2024

Published online: 29 January 2024

## References

- Bachman, L., Davidson, F., Ryan, K., & Choi, I. (1995). *An investigation into the comparability of two tests of English as a foreign language: The Cambridge-TOEFL comparability study*. Cambridge University Press.
- Byrne, B. M. (2016). *Structural equation modeling with Amos: Basic concepts, applications, and programming* (3rd ed.). Routledge.
- Committee for Selection of Good Practices in University Admissions. (2022). *Collection of good practices in university admissions*. Higher Education Bureau MEXT.
- Council of Europe. (2020). *Common European Framework of References for Languages: Learning, teaching, and assessment: Companion volume*. Council of Europe. Retrieved September 25 from <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Council of Local Authorities for International Relations. (2015). *Sankakoku*. Council of Local Authorities for International Relations. Retrieved December, 27 from <https://jetprogramme.org/ja/countries/>
- Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement*, 28(4), 227–246. <https://doi.org/10.1177/0146621604265031>
- Educational Testing Service. (2020). *Practice test: TOEIC Bridge listening & reading test*. Educational Testing Service. Retrieved December, 27 from [https://etswebsiteprod.cdn.prismic.io/etswebsiteprod/c2be94a8-aea3-45a4-a742-2b83bcff8960\\_The+Redesigned+TOEIC+Bridge+Test+-+Sample+test.pdf](https://etswebsiteprod.cdn.prismic.io/etswebsiteprod/c2be94a8-aea3-45a4-a742-2b83bcff8960_The+Redesigned+TOEIC+Bridge+Test+-+Sample+test.pdf)
- Eiken Foundation of Japan. (2015). *Daigakunyushi centersiken tonou soukanchousa: Jitsuyoueigo ginoukentei to TEAP de jissi*. [https://www.eiken.or.jp/teap/info/2015/pdf/20151007\\_pressrelease\\_testresearch.pdf](https://www.eiken.or.jp/teap/info/2015/pdf/20151007_pressrelease_testresearch.pdf)
- Everson, P., Duke, T., Gomez, P. G., Grissom, C. E., Park, E., & Schmidgall, J. (2021). Development of the redesigned TOEIC Bridge® Tests: A compendium of studies. In J. Schmidgall (Ed.), *The research foundation for the redesigned TOEIC Bridge® tests*. (Vol. 4). ETS.
- Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble. *European Journal of Psychological Assessment*, 33(6), 399–402. <https://doi.org/10.1027/1015-5759/a000460>
- Fouly, K. A., Bachman, L. F., & Cziko, G. A. (1990). The divisibility of language competence: A confirmatory approach. *Language Learning*, 40(1), 1–21. <https://doi.org/10.1111/j.1467-1770.1990.tb00952.x>
- Gu, L. (2015). Language ability of young English language learners: Definition, configuration, and implications. *Language Testing*, 32(1), 21–38. <https://doi.org/10.1177/0265532214542670>
- Hale, G. A., Stansfield, C. W., Rock, D. A., Hicks, M. M., Butler, F. A., & Oller, J. W. (1988). Multiple-choice cloze items and the Test of English as a Foreign Language. *ETS Research Report Series*, (1), i-117. <https://doi.org/10.1002/j.2330-8516.1988.tb00258.x>
- Hale, G. A., Rock, D. A., & Jirele, T. (1989). Confirmatory factor analysis of the Test of English as a Foreign Language. *ETS Research Report Series*, (2), i-51. <https://doi.org/10.1002/j.2333-8504.1982.tb01327.x>
- IIBC. (2021). *Comparative analysis of scores between the Common Test for University Admissions and TOEIC Program*.
- IIBC. (2022). *Comparative analysis of scores between the Common Test for University Admissions and TOEIC Program*.

- IIBC. (n.d.). *Sample mondai*. IIBC. Retrieved March 27, 2023 from <https://www.iibc-global.org/toeic/pr/bridge4skills/sample.html>
- In'nami, Y., Koizumi, R., & Nakamura, K. (2016). Factor structure of the Test of English for Academic Purposes (TEAP) test in relation to the TOEFL iBT test. *Language Testing in Asia*, 6(1). <https://doi.org/10.1186/s40468-016-0025-9>
- In'nami, Y., & Koizumi, R. (2012). Factor structure of the revised TOEIC test: A multiple-sample analysis. *Language Testing*, 29(1), 131–152. <https://doi.org/10.1177/0265532211413444>
- Kamiya, N. (2017). Can the National Center Test in Japan be replaced by commercially available private English tests of four skills? In the case of TOEFL Junior Comprehensive. *Language Testing in Asia*, 7(1). <https://doi.org/10.1186/s40468-017-0046-z>
- Kawaijuku Education Institution. (n.d.). *Eigo kentei tokutei shiken wo riyousuru daigaku ichiran hyouno mikata*. Kawaijuku Education Institution. Retrieved December, 24 from [https://www.keinet.ne.jp/exam/2023/pdf/23eigo\\_shikaku.pdf](https://www.keinet.ne.jp/exam/2023/pdf/23eigo_shikaku.pdf)
- Kunnan, A. (1995). *Test taker characteristics and test performance*. Cambridge University Press.
- Livingston, S. A. (2014). *Equating test scores (without IRT)*. Educational Testing Service. <https://www.ets.org/Media/Research/pdf/LIVINGSTON2ed.pdf>
- Manning, W. H. (1987). Development of cloze-elide tests of English as a second language. *ETS Research Report Series*, (1), i-136. <https://doi.org/10.1002/j.2330-8516.1987.tb00222.x>
- MEXT. (2018a). *The government course guidelines for senior high school*.
- MEXT. (2018b). *Koutou gakkou gakushuu shidou yoryo kaisetsu gaikokugo eigo hen*. Ministry of Education, Culture, Sports, Science and Technology. Retrieved August 31 from [https://www.mext.go.jp/content/1407073\\_09\\_1\\_2.pdf](https://www.mext.go.jp/content/1407073_09_1_2.pdf)
- MEXT. (2018c). *Summary of English proficiency survey results (Third-year high school students)*. [https://www.mext.go.jp/a\\_menu/kokusai/gaikokugo/\\_icsFiles/afieldfile/2018/04/06/1403470\\_03\\_1.pdf](https://www.mext.go.jp/a_menu/kokusai/gaikokugo/_icsFiles/afieldfile/2018/04/06/1403470_03_1.pdf)
- Mitsumura Toshio. (n.d.). *Kyokasho dewa America no eigo to Igrisu no eigo no dochirawo tsukatte iruno desuka*. Mitsumura Toshio. Retrieved December, 27 from <https://www.mitsumura-toshio.co.jp/webmaga/kotoba-to-manabi/eigo-qa/detail03>
- Mizumoto, A. (2014). Sokuteino datouseito shinraisei -yuidatano hissujuyouken towa-. In O. Takeuchi & A. Mizumoto (Eds.), *Gaikokugo kyouiku kenkyu handbook-kaiteiban-* (pp. 17–31). Shohakusha.
- Nakamura, K. (2022). Investigating the factor structure of the Test of English for Academic Purposes (TEAP) and its relation to test takers' perceived test task value. *Language Testing in Asia*, 12(1). <https://doi.org/10.1186/s40468-022-00183-4>
- National Center for University Entrance Examinations. (2021). *Reiwa 3 nendo daigaku nyugaku kyoutsu testo eigo ni tsuite*. National Center for University Entrance Examinations. Retrieved December, 27 from [https://www.mext.go.jp/content/20210419-mxt\\_daigakuc02-000014254\\_3\\_1.pdf](https://www.mext.go.jp/content/20210419-mxt_daigakuc02-000014254_3_1.pdf)
- National Center for University Entrance Examinations. (n.d.-a). *Daigaku kyoutsu testo eigo ni okeru Igrisu eigo no shiyou ni tsuite*. National Center for University Entrance Examinations. Retrieved December, 27 from [https://www.dnc.ac.jp/albums/abm.php?d=58&f=abm00000705.pdf&n=R4\\_大学入学共通テスト英語におけるイギリス英語の使用について.pdf](https://www.dnc.ac.jp/albums/abm.php?d=58&f=abm00000705.pdf&n=R4_大学入学共通テスト英語におけるイギリス英語の使用について.pdf)
- National Center for University Entrance Examinations. (n.d.-b). *Reiwa 5 nendo daigaku nyugakusha senbatsu ni kakaru daigaku nyugaku kyoutsu tesuto mondai sakusai houshin*. National Center for University Entrance Examinations. Retrieved December, 27 from <https://www.dnc.ac.jp/albums/abm.php?d=494&f=abm00000288.pdf&n=令和5年度大学入学者選抜に係る大学入学共通テスト出題教科・科目の出題方法及び問題作成方針.pdf>
- National Center for University Entrance Examinations. (n.d.-c). *Role of Common Test*. [https://www.dnc.ac.jp/kyoutsu/shiken\\_gaiyou/yakuwari.html](https://www.dnc.ac.jp/kyoutsu/shiken_gaiyou/yakuwari.html)
- National Center for University Entrance Examinations. (n.d.-d). *Summary of results of Common Test*. National Center for University Entrance Examinations. Retrieved March 21, 2023 from <https://www.dnc.ac.jp/studentandparent/>
- O'Sullivan, B., Motteram, J., Skipsey, R., & Dunlea, J. (2022). *The importance of the four skills in the Japanese context*. British Council. Retrieved December, 26 from [https://www.britishcouncil.jp/sites/default/files/importance\\_of\\_four\\_skills\\_japanese\\_context\\_rev\\_0.pdf](https://www.britishcouncil.jp/sites/default/files/importance_of_four_skills_japanese_context_rev_0.pdf)
- Oller, J. W. J. (1979). *Language tests at school: A pragmatic approach*. Longman.
- Oller, J. W., & Hinofotis, F. B. (1980). Two mutually exclusive hypotheses about second language ability: Indivisible or partially divisible competence. In J. W. J. Oller & K. Perkins (Eds.), *Research in language testing* (pp. 13–23). Newbury House.
- Otsu, T. (2014). *Hyoujiyunka eigoshiken to center shiken eigokamoku tokuten tono kankei bunseki*. [http://antlers.rd.dnc.ac.jp/~otsu/doc/jart2014\\_p66-69.pdf](http://antlers.rd.dnc.ac.jp/~otsu/doc/jart2014_p66-69.pdf)
- Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Sawaki, Y., & Sinharay, S. (2018). Do the TOEFL iBT® section scores provide value-added information to stakeholders? *Language Testing*, 35(4), 529–556. <https://doi.org/10.1177/0265532217716731>
- Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing*, 26(1), 005–030. <https://doi.org/10.1177/0265532208097335>
- Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2008). *Factor structure of the TOEFL Internet-based test (iBT): Exploration in a field trial sample* (TOEFL iBT Research Report, Issue). Educational Testing Service. <http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.2008.tb02095.x/epdf>
- Schmidgall, J., Oliveri, M. E., Duke, T., & Grissom, C. E. (2019). Justifying the construct definition for a new language proficiency assessment: The redesigned TOEIC Bridge tests—Framework paper. *ETS Research Report Series*, 2019(1), 1–20. <https://doi.org/10.1002/ets2.12267>
- Schmidgall, J., Cid, J., Grissom, C. E., & Li, L. (2021). Making the case for the quality and use of a new language proficiency assessment: Validity argument for the redesigned TOEIC Bridge tests. *ETS Research Report Series*, 2021(1), 1–22. <https://doi.org/10.1002/ets2.12335>
- Schmidgall, J. (2021). Mapping the redesigned TOEIC Bridge test scores to proficiency levels of the Common European Framework of Reference for Languages. *ETS Research Memorandum Series*. <https://files.eric.ed.gov/fulltext/ED617522.pdf>
- Shin, S.-K. (2005). Did they take the same test? Examinee language proficiency and the structure of language tests. *Language Testing*, 22(1), 31–57. <https://doi.org/10.1191/0265532205lt296oa>

- Stricker, L. J., & Rock, D. A. (2008). *Factor structure of the TOEFL Internet-Based Test across subgroups* (TOEFL iBT Research Report, Issue). Educational Testing Service. <https://www.ets.org/Media/Research/pdf/RR-08-66.pdf>
- Swinton, S. S., & Powers, D. E. (1980). Factor analysis of the Test of English as a Foreign Language for several language groups. *ETS Research Report Series*, (2), i-79. <https://doi.org/10.1002/j.2333-8504.1980.tb01229.x>
- Tergujeff, E. (2023). Second language learners listening to their peers: Is there a shared L1 effect for L2 comprehensibility and accentedness? *European Journal of Applied Linguistics*. <https://doi.org/10.1515/eujal-2022-0014>
- Tokyo Shoseki. (2018). *Kyokasho tosho kyozai yokuaru goshitsumon Q&A: Chugakko Eigo*. Tokyo Shoseki. Retrieved December, 27 from <https://www.tokyo-shoseki.co.jp/question/j/eigo.html>
- Wilson, K. M. (2000). An exploratory dimensionality assessment of the TOEIC test. *ETS Research Report Series*, 2000(2), i-28. <https://doi.org/10.1002/j.2333-8504.2000.tb01837.x>

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.