An Investigation of IELTS and TOEFL iBT Score Compatibility

Masaya Kanzaki
Kanda University of International Studies

Reference Data:

Kanzaki, M. (2024). An investigation of IELTS and TOEFL iBT score compatibility. In B. Lacy, R. P. Lege, & P. Ferguson (Eds.), *Growth Mindset in Language Education*. JALT.

This study details an analysis of scores from 53 university students in Japan on the International English Language Testing System (IELTS) test and the Test of English as a Foreign Language Internet-Based Test (TOEFL iBT) and assesses the validity of the two score comparison tables published by the Educational Testing Service (2010) and by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT, 2018). The findings indicate that 27 participants (50.9%) obtained the corresponding scores suggested in the Educational Testing Service table, and only 15 (28.3%) obtained the CEFR-based corresponding scores suggested in the Ministry of Education table. Additionally, it was observed that participants who had more experience with either of the tests tended to score higher on that test than what was projected in the Educational Testing Service table.

本研究は、53名の日本の学生から集めたInternational English Language Testing System（IELTS）とTest of English as a Foreign Language Internet-Based Test（TOEFL iBT）の得点を分析し、Educational Testing Service（2010）と文部科学省（2018）の得点対照表の妥当性を検証した。分析の結果、27名（50.9%）の参加者がEducational Testing Serviceの表で示されている対応得点を取得したが、文部科学省の表が示すCEFRに基づく対応得点を取ったのは、15名（28.3%）にとどまった。さらに、いずれかの試験の受験経験が豊富な参加者は、その試験の得点が、Educational Testing Serviceの表で予測されている対応得点よりも高くなる傾向があることが示された。

https://doi.org/10.37546/JALTPCP2023-43

This article was originally published in the Selected Papers section of the 2023 Postconference Publication (PCP), *Growth Mindset in Language Education*. The PCP publishes papers based on presentations given at the JALT International Conference, and the Selected Papers section highlights a small number of papers of exceptional quality that have been first suggested by the editorial staff and then vetted by the JALTPublications Board through a blind review process. We feel that papers like this one represent some of the best work that the JALT Conference and the PCP have to offer, and encourage interested readers to check out other selected papers at https://jalt-publications.org/proceedings.

This small-scale study compares scores on the paper-based International English Language Testing System (IELTS) test and the Test of English as a Foreign Language Internet-Based Test (TOEFL iBT), two of the most widely accepted English-proficiency tests in the world. English-medium universities worldwide often include scores of both tests in their English language requirements for applicants from outside English-speaking countries. In Japan, an increasing number of Japanese universities have incorporated commercially available English tests into their selection processes for admission. This shift is in response to a recommendation by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT, 2021a), and the IELTS test and TOEFL iBT are now among the accepted tests at a number of Japanese universities (Kawaijuku, 2023).

As many universities around the world accept both tests, examining the score linkage between them would be of interest to such universities and their applicants. Davies et al. (1999), however, challenged the concept of test equivalence:

> Strictly speaking, this concept is unjustifiable, since each test is designed for a different purpose and a different population, and may view and assess language traits in different ways as well as describing test-taker performance differently. (p. 199)

Even so, they acknowledged that test users may demand statements of equivalence between different tests and suggested that the demand can be met by conducting a concurrent validity study, in which scores on different tests taken by the same group of people are compared. Test providers have carried out such studies and produced score comparison tables between their own and their competitors' tests (e.g., Cardwell et al., 2024; Clesham & Hughes, 2020; Educational Testing Service [ETS], 2010; Elliot et al., 2021; Kim et al., 2017).

Similar to those studies, this study reviewed the scores of different tests taken by the same group.

However, due to the limited sample size, the intention of the study was not to create a score comparison table. Instead, it provided evaluation of the accuracy of two existing comparison tables, one published by ETS (2010) and the other by MEXT (2018). The ETS table is grounded in the results of a concurrent validity study, but the score linkages in the MEXT table are based on the Common European Framework of Reference for Languages (CEFR) levels, a difference that raises the question of their relative accuracy. Thus, the purpose of this study was to examine the relationship between IELTS and TOEFL scores and to evaluate the accuracy of the ETS and MEXT tables.

Method

Participants

The participants in this study were 53 students (36 females and 17 males), including one postgraduate student from an MA TESOL program and 52 undergraduates (four in their 1st year, 10 in their 2nd, 21 in their 3rd, and 17 in their 4th). All were attending a university emphasizing foreign language education, and as their main foreign language, 41 were studying English, while 11 were studying other languages, such as Spanish, Chinese, and Vietnamese. All but four were first-language (L1) Japanese speakers, and these four were L1 speakers of Chinese, French, and Thai but also proficient in Japanese.

Participants were recruited via the university's web portal. Incentives for participation included fee waivers for the two tests and an Amazon coupon worth 5,000 yen. The research grant allowed for the participation of only 56 individuals, but 151 students applied, and so selections were made on the basis of TOEIC L&R or TOEFL ITP scores. Since the IELTS test and TOEFL iBT are designed to assess English language ability needed for higher education and are not suitable for low-proficiency students, only those with TOEIC scores of 730 and above or TOEFL ITP scores of 513 and above were chosen. Ethical approval was obtained from the university's institutional review board, and all the participants provided informed consent. The study started with 56 participants, but one missed the IELTS test and two missed the TOEFL iBT and were therefore excluded.

Procedures

All participants took the IELTS test and TOEFL iBT in a designated testing center between July and November 2022. They were instructed to select exam dates within the window of August to September 2022 and to take both tests within a month of each other. However, three participants had to reschedule their exams due to illness or transportation issues. Another accidentally cancelled her score at the end of the TOEFL iBT, so she submitted the score report for the test she had taken in July. The scores of these four participants were included in the study, even though the two tests had been taken more than a month apart, on the ground that English proficiency levels do not generally change much in such a short time. Out of the 53 participants in total, 28 took the IELTS test first, and 25 took the TOEFL iBT first. (For the exam dates of each participant, see the webpage for this study; https://bit.ly/pcp2023mk.) Descriptive statistics for each group of scores as well as correlations between them were then calculated, and the ETS and MEXT tables were evaluated for accuracy.

Materials

IELTS and TOEFL iBT

The paper-based IELTS Academic test and the computer-based TOEFL iBT were used in the study. Note that the TOEFL iBT has been updated multiple times since its inception in 2005 and that the version used in this study is different from that used in ETS (2010) and from the one most recently introduced in July 2023.

Table 1 compares the components of the two tests. For more details on the IELTS test, see Read (2022), and for more details on the version of the TOEFL iBT used in this study, see ETS (2020a).

Table 1

Components of the IELTS Test and the TOEFL iBT

	IELTS	TOEFL
Listening	2 conversations and 2 monologues, 40 questions, 30 minutes	2–3 conversations and 3–4 monologues, 28–39 questions, 41–57 minutes
Reading	3 passages, 40 questions, 60 minutes	3–4 passages, 30–40 questions, 54–72 minutes
Speaking	3 parts, 11–14 minutes	4 tasks, 17 minutes
Writing	2 tasks (150 words and 250 words), 60 minutes	2 tasks (150–225 words and 300 words), 50 minutes

The IELTS section and overall scores range from 0 to 9 with a half-point interval. The overall score is the average of the four section scores rounded to the nearest half-point. The TOEFL section scores range from 0 to 30 with a full-point interval, and the total score is the sum of the four section scores, ranging from 0 to 120.

Reliability estimates for the IELTS test are .91 for the listening scores, .90 for the reading, .87 for the speaking, .92 for the writing, and .97 for the overall scores (IELTS, 2023). Reliability estimates for the TOEFL iBT are .87 for the listening scores, .87 for the reading, .86 for the speaking, .80 for the writing, and .95 for the total scores (ETS, 2020b).

ETS Table

ETS (2010) published five score comparison tables linking section and overall/total scores of the IELTS test and TOEFL iBT. The tables were derived from the scores of 1,153 test takers from 70 countries, including China (41%), the United States (6%), and Japan (5%). The score data were collected in 2008 and 2009. The scores were linked by way of the equipercentile linking method, in which scores are equated based on their percentile ranks. Because the score scales of the two tests are different, the IELTS scores were linked to ranges of TOEFL scores.

Among the five tables, this study evaluated the one that links the IELTS overall and TOEFL total scores. Table 2 shows part of the correspondence in the table. IELTS scores of 7.5 and above and their corresponding TOEFL score ranges are excluded. Note that the TOEFL score ranges are not uniform. For example, the TOEFL score range for an IELTS score of 6 covers 19 scores (between 60 and 78), but the score range for an IELTS of 5 only covers 11 scores (between 35 and 45).

MEXT Table

MEXT (2018) published a comparison table linking scores from eight English proficiency tests (Cambridge English Qualifications, Eiken, GTEC, IELTS, TEAP, TEAP CBT, TOEFL iBT, and TOEC L&R+S&W) with the six CEFR levels, ranging from A1 (lowest) to C2 (highest). These linkages were established on the basis of the test providers' assessments of the relationships between their respective tests and the CEFR levels. The table was created as part of MEXT's effort to integrate these eight tests into the Japanese university entrance selection system. Although the plan for the new system was abandoned in 2019, some universities have incorporated commercially available tests into their selection processes, and the table continues to hold influence among them (MEXT, 2021b).

Table 3 shows the CEFR-linked correspondence of the MEXT table in the left half and a subdivided version of the correspondence in the right half. Because the ETS table links one IELTS score to a range of TOEFL scores, a subdivided version was created to include the comparable divisions by dividing each CEFR level of the original correspondence by 3. At the B1 level of the subdivided version, IELTS scores of 4, 4.5, and 5 are linked to 10 TOEFL scores each. At the B2 level, IELTS scores of 5.5 and 6.5 are linked to eight TOEFL scores each and IELTS scores of 6 to seven TOEFL scores. At the C1 level, IELTS scores of 7 and 8 are linked to nine TOEFL scores each and IELTS scores of 7.5 to eight TOEFL scores.

Table 3

Correspondence Between the IELTS and TOEFL Scores in the MEXT Table

CEFR level	Original linkage		Each level divided by 3	
	IELTS	TOEFL	IELTS	TOEFL
C1	7–8	95–120	8	112–120
			7.5	104–111
			7	95–103
B2	5.5–6.5	72–94	6.5	87–94
			6	80–86
			5.5	72–79
B1	4–5	42–71	5	62–71
			4.5	52–61
			4	42–51

Table 4 compares the TOEFL score ranges of the ETS table (Table 2), and the subdivided version of the MEXT table, shown in the right half of Table 3.

Table 2

Correspondence Between the IELTS and TOEFL Scores in the ETS Table

IELTS	0–4	4.5	5	5.5	6	6.5	7
TOEFL	0–31	32–34	35–45	46–59	60–78	79–93	94–101

The TOEFL score ranges of the two tables corresponding to IELTS scores of 7 are similar, and there is a significant overlap in the score ranges corresponding to IELTS scores of 6.5. However, for IELTS scores of 6 and below, there are no overlaps in the TOEFL score ranges between the two tables.

Results

Descriptive Statistics

Table 5 presents the descriptive statistics for the section and overall/total scores of the IELTS test and TOEFL iBT. The observed score ranges and the standard deviations indicate that the listening and reading scores are more widely spread out than the speaking and writing scores on both tests.

Among the four section mean scores of the IELTS test, the reading score is the highest, followed by the listening and writing scores, and the speaking score is the lowest. Among the four section mean scores of the TOEFL iBT, the writing score is the highest, followed by the listening and speaking scores, and the reading score is the lowest. The IELTS mean scores suggest that reading is the participants' strongest skill, but the TOEFL mean scores suggest that it is their weakest. This finding may indicate that the two tests assess and score English abilities in different ways, as suggested by Davies et al. (1999).

Correlations

The results of a Kolmogorov-Smirnov test indicate that seven out of 10 sets of scores do not follow a normal distribution, and the histograms show that all 10 sets of scores deviate from a normal distribution. When the assumption of normality is violated, nonparametric statistics, such as Spearman's r_s correlation, should be used to analyze the data. However, it is a convention of score comparison studies to report Pearson's *r* correlations, so both types of correlations were calculated between the section and overall/total scores of the two tests. For non-normally distributed data sets, bootstrap methods are recommended to generate confidence intervals (Field, 2024; Plonsky et al., 2013; Puth et al., 2015), so bootstrapping was performed with 2,000 samples, and bias-corrected and accelerated bootstrap 95% confidence intervals were obtained

Table 4

Comparison of Correspondences in the ETS and Subdivided MEXT Tables

IELTS	0–4*/4	4.5	5	5.5	6	6.5	7
TOEFL (ETS)	0–31	32–34	35–45	46–59	60–78	79–93	94–101
TOEFL (MEXT)	42–51	52–61	62–71	72–79	80–86	87–94	95–103

Note. *The ETS table links IELTS 0–4 to TOEFL 0–31.

Table 5

Descriptive Statistics for Scores

Test	**Score type**	**Possible score**	**Score range**	***SD***	***M***
IELTS	Listening	0–9	4.5–7.5	0.8	5.9
	Reading	0–9	4–8.5	0.8	6.0
	Speaking	0–9	4–7	0.6	5.6
	Writing	0–9	5–7	0.5	5.7
	Overall	0–9	5–7	0.5	5.9
TOEFL	Listening	0–30	5–28	4.9	16.2
	Reading	0–30	3–25	4.6	15.9
	Speaking	0–30	6–22	3.2	16.0
	Writing	0–30	10–21	3.1	16.4
	Total	0–120	30–88	13.0	64.6

Note. $N = 53$.

for all correlations. All correlations were significant and are displayed in Table 6.

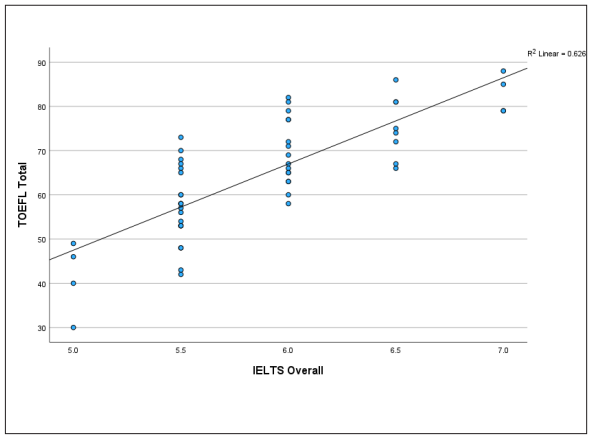
Comparing the two types of correlations, they are the same for the listening and overall/total scores, the Spearman's r_s correlations are higher for the reading and writing scores, and the Pearson's *r* correlation is higher for the speaking scores. Overall, the two types of correlations are similar.

ETS (2010) reported that the listening, reading, speaking, writing, and overall/total score correlations between the two tests were .63, .68, .57, .44, and .73, respectively. Compared to these, the listening, reading, and speaking score correlations are lower in the present study, but the writing and overall/total score correlations are higher.

Figure 1 shows a scatterplot that visually presents the relationship between the IELTS overall and TOEFL total scores. Because all the participants received one of the five IELTS scores between 5 and 7, the dots are aligned on the five vertical lines representing the five scores.

Figure 1

Scatterplot of the IELTS Overall and TOEFL Total Scores (N = 53)



Evaluating the ETS and MEXT Tables

Table 7 uses the ETS table grid and shows how many participants obtained what combination of an IELTS overall score and a TOEFL total score range. For example, "12" in the cell where the column for IELTS 6 and the row for TOEFL 60–78 intersect means that 12 participants received an IELTS score of 6 and a TOEFL score between 60 and 78. Shaded cells indicate that the IELTS scores and TOEFL score ranges align with the correspondence in the ETS table.

The correspondence rate was highest among for participants with IELTS scores of 6 (12 out of 16, 75%), followed by IELTS scores of 5.5 (11 out of 21, 52.4%). The rates are lower for those with IELTS scores of 5 (one out of four, 25%) and 6.5 (three out of eight, 37.5%), and none of the four participants with an IELTS score of 7 obtained a TOEFL score within the corresponding range. In total, 27 out of 53 (50.9%) obtained the published corresponding scores.

Among participants with IELTS scores of 5 and 5.5, 10 out of 25 (40%) obtained a TOEFL score above the predicted corresponding ranges. However, among those with IELTS scores of 6.5 and 7, none obtained a TOEFL score above the predicted corresponding ranges.

Table 8 uses the MEXT table grid and shows the numbers of participants obtaining each combination scores on the two tests. The areas outlined in boldface indicate where the same CEFR level scores of the two tests meet, while the shaded cells indicate where the corresponding scores in the subdivided version of the table meet.

For the CEFR level correspondence, two out of four participants (50%) with B1 level IELTS scores and 13 out of 45 (28.9%) with B2 level IELTS scores received the same CEFR level TOEFL scores. In total, 15 out of 53 (28.3%) received the predicted corresponding CEFR level scores. The TOEFL scores of the remaining 38 (71.7%) were one level below the CEFR levels linked to their IELTS scores. It seems that the predicted B2 level IELTS scores are equiva-

Table 6

Correlations Between the IELTS and TOEFL Scores

	Listening	Reading	Speaking	Writing	Overall/Total
Spearman's r_s	.56*	.57*	.40***	.61*	.79*
	[.38, .71]	[.35, .74]	[.14, .60]	[.44, .74]	[.67, .88]
Pearson's *r*	.56*	.54*	.42**	.57*	.79*
	[.42, .69]	[.33, .74]	[.22, .59]	[.42, .72]	[.70, .87]

Note. $N = 53$. *$p < .001$ (two-tailed). **$p = .002$ (two-tailed). ***$p = .003$ (two-tailed). BCa bootstrap 95% CIs reported in brackets.

lent to the B1 level TOEFL scores. For the subdivided correspondence, only three out of 53 (5.7%) test takers obtained the predicted corresponding scores.

Table 7

Relationship Between the IELTS and TOEFL Scores With the ETS Table Grid

		IELTS				
		5	**5.5**	**6**	**6.5**	**7**
TOEFL	94–101					
	79–93			3	3	4
	60–78		8	12	5	
	46–59	2	11	1		
	35–45	1	2			
	32–34					
	0–31	1				

Note. $N = 53$.

Table 8

Relationship Between the IELTS and TOEFL Scores With the MEXT Table Grid

			IELTS				
			B1	**B2**			**C1**
			5	**5.5**	**6**	**6.5**	**7**
TOEFL	**C1**	95–103					
	B2	87–94					
		80–86			2	3	2
		72–79		1	4	3	2
	B1	62–71		5	8	2	
		52–61		11	2		
		42–51	2	4			
	NA	Below 42	2				

Note. $N = 53$.

Table 9 shows the average TOEFL scores of the five IELTS score groups. Among the five TOEFL average scores, those of the IELTS 5, 5.5, and 6 groups are within the predicted corresponding ETS table score ranges, whereas none of the five scores are within the predicted corresponding score ranges of the subdivided MEXT table.

Effects of Previous Test-Taking Experience

Of the 53 study participants, 35 were taking the IELTS test and the TOEFL iBT for the first time. Among the 18 who had taken either or both tests before the study, seven had taken the IELTS test only, five had taken the TOEFL iBT only, and six had taken both. Among the six who had taken both, one had taken the IELTS test more times, three had taken the TOEFL iBT more times, and two had taken both tests an equal number of times. Altogether, 37 had taken both tests an equal number of times (including those who had not taken either), eight had taken the IELTS test more times, and eight had taken the TOEFL iBT more times.

Table 10 shows the relationship between the participants' test-taking experience and their scores in relation to the correspondence to the ETS table. The participants have been divided into three groups: those who had taken the TOEFL iBT more times, those who had taken both tests an equal number of times (including those who had not taken either), and those who had taken the IELTS test more times. Their TOEFL iBT total scores are divided into three categories in relation to the corresponding score ranges: above, within, and below the supposed corresponding score ranges.

Among those who had taken the TOEFL iBT more times, five out of eight (62.5%) obtained a TOEFL score above their predicted corresponding score range. Among those who had taken both tests an equal number of times (including those who had not taken either), 22 out of 37 (59.5%) obtained a TOEFL score within their predicted corresponding score range. Among those who had taken the IELTS

Table 9

Average TOEFL Scores of the Five IELTS Score Groups

IELTS group	**5**	**5.5**	**6**	**6.5**	**7**
TOEFL average	41.3	57.6	69.7	75.3	82.8
TOEFL (ETS)	35–45	46–59	60–78	79–93	94–101
TOEFL (MEXT)	62–71	72–79	80–86	87–94	95–103
n	4	21	16	8	4

Note. $N = 53$.

test more times, five out of eight (62.5%) obtained a TOEFL score that was below their predicted corresponding score range.

Table 10

Relationship Between Test-Taking Experience and Scores

Test-taking experience	TOEFL score in relation to the corresponding range		
	Above	Within	Below
More TOEFL	5	2	1
Equal times	8	22	7
More IELTS	0	3	5

Note. $N = 53$.

Discussion

The correlation between the IELTS overall and TOEFL total scores found in this study was .79, which is moderately high but arguably not high enough for scores of the two tests to be used interchangeably. For scores of different tests to be interchangeable, a certain level of correlation is required, and Dorans (2000, 2004) argued that the threshold for credible interchangeability is .866, based on the rate of reduction in uncertainty. However, correlations between scores of different English proficiency tests are rarely in the high .80s and typically below .80. Examples of score correlations between major English proficiency tests include .73 between the IELTS test and TOEFL iBT (ETS, 2010), .74 between the IELTS test and Pearson Test of English Academic (Clesham & Hughes, 2020), .71 between the TOEFL iBT and Duolingo English Test (DET), and .65 between the IELTS test and DET (Cardwell et al., 2023). Even so, universities that use IELTS and TOEFL scores interchangeably should be aware that their practice of doing so could be considered psychometrically unsound.

This lower-than-ideal correlation may have contributed to the low accuracy rates of the ETS and MEXT tables; 27 out of 53 (50.9%) obtained the corresponding scores in the ETS table, and 15 out of 53 (28.3%) obtained the corresponding CEFR level scores in the MEXT table. Another factor that may have affected the accuracy of the correspondences was the participants' prior experience with the two tests: Those who had taken the IELTS test more times than the TOEFL iBT were more likely to get a higher IELTS score relative to their TOEFL score, and vice versa. Among those who had taken both tests an equal number of times (including those who had not taken either), the accuracy rate of the ETS table was higher (22 out of 37, 59.5%).

Conclusion

An important implication of this study is that the IELTS and TOEFL score correspondence in the MEXT table is much less than ideal and, therefore, Japanese universities should not rely on it. For example, a number of Japanese universities treat applicants with an IELTS score of 4 and those with a TOEFL iBT score of 42 equally in their selection processes (Kawaijuku, 2023) because both scores are linked in the MEXT table as the lowest of the CEFR B1 level scores. However, according to the findings of this study, an IELTS score of 4 seems to be equivalent to a range of TOEFL iBT scores around 30, and a TOEFL iBT score of 42 is equivalent to an IELTS score of 5.

Another implication related to the accuracy of a score comparison table is that it is difficult to link scores of two English tests accurately, even when a concurrent validity study is conducted. This difficulty stems from differerences in how each test is designed, how it views and assesses language abilities, and how it describes test-taker performance, as Davies et al. (1999) suggested.

A practical implication for learners is that if they take the same English test repeatedly, they might get relatively better results on that test than on another with which they have little experience. As such, if they fail to obtain a required score on one test, it would be better to try the same test again than switch to another.

A limitation of this study is its small sample size, which warrants caution in considering the results. For future research, a study with a bigger sample would be desirable. In fact, ETS has sent out emails calling for volunteers to participate in a new score comparison study between the IELTS test and TOEFL iBT. It would be interesting to see how the new comparison table will be different from the one examined in this study.

Data Sharing

The score data, SPSS output, exam dates, and other information related to this study are available at https://bit.ly/pcp2023mk.

Acknowledgments

This study was supported by JSPS KAKENHI Grant Number 20K00891.

References

Cardwell, R. L., Naismith, B., LaFlair, G. T., & Nydick, S. W. (2023). *Duolingo English Test: Technical manual* (Duolingo Research Report). Retrieved April 7, 2024, from https://duolingo-papers.s3.amazonaws.com/other/technical_manual.pdf

Cardwell, R. L., Nydick, S. W., Lockwood, J. R., & von Davier, A. A. (2024). Practical considerations when building concordances between English tests. *Language Testing 41*(1), 192–202. https://doi.org/10.1177/02655322231195027

Clesham, R., & Hughes, S. R. (2020). *2020 concordance report PTE Academic and IELTS Academic*. Retrieved April 7, 2024, from https://www.pearsonpte.com/ctf-assets/yqwtwibiobs4/1hXHbkTLYCJly7JryACWjK/5a20dbe26d8ca2c36a3b0dd5a32868d7/2021_PTEA_2020_PTE_IELTS_Concordance_White_Paper.pdf

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing* (Vol. 7). Cambridge University Press.

Dorans, N. J. (2000). *Distinctions among classes of linkages* (College Board Research Note RN-11). 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. (2010). *Linking TOEFL iBT scores to IELTS scores – A research report*. Retrieved from https://www.ets.org/pdfs/toefl/linking-toefl-ibt-scores-to-ielts-scores.pdf

Educational Testing Service. (2020a). TOEFL iBT test framework and test development. *TOEFL Research Insight Series* (Vol. 1, 3rd ed.). Retrieved from https://www.ets.org/pdfs/toefl/toefl-ibt-insight-s1v1.pdf

Educational Testing Service. (2020b). Reliability and comparability of TOEFL iBT scores. *TOEFL Research Insight Series* (Vol. 3, 3rd ed.). Retrieved from https://www.ets.org/pdfs/toefl/toefl-ibt-insight-s1v3.pdf

Elliot, M., Blackhurst, A., O'Sullivan, B., Clark, T., Dunlea, J., & Saville, N. (2021). Aligning IELTS and PTE-Academic: A measurement study. In N. Saville, B. O'Sullivan, & T. Clark (Eds.), *IELTS Partnership Research Papers: Studies in Test Comparability Series,* No. 2, (pp. 42–63). IELTS Partners. Retrieved April 7, 2024, from https://takeielts.britishcouncil.org/sites/default/files/relationship_ielts_and_pte_academictc.pdf

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th ed.). Sage Publications Limited.

IELTS. (2023). *Test performance 2022* [Web page]. Retrieved March 10, 2024, from https://ielts.org/researchers/our-research/test-statistics

Kawaijuku. (2023). Eigo shikaku-, kentei shiken-o riyō suru daigaku ichiran (英語資格・検定試験を利用する大学一覧) [A list of universities that use English proficiency tests for selection processes]. Kawaijuku. Retrieved March 10, 2024, from https://www.keinet.ne.jp/exam/2024/pdf/24eigo_shikaku.pdf

Kim, M., Smith, W. Z., & Chin, T. Y. (2017). *Validation and linking scores for the Global Test of English Communication: White paper*. Retrieved from https://www.benesse.co.jp/gtec/schoolofficials/research/pdf/doc-2016-02.pdf

Ministry of Education, Culture, Sports, Science and Technology–Japan. (2018). Kakushikaku-, kentei shiken-to CEFR-to-no taishōhyō (各資格・検定試験とCEFRとの対照表) [Comparison table of English proficiency tests and the CEFR levels]. MEXT. Retrieved from https://www.mext.go.jp/b_menu/shingi/chousa/koutou/091/gijiroku/__icsFiles/afieldfile/2018/07/27/1407616_003.pdf

Ministry of Education, Culture, Sports, Science and Technology–Japan. (2021a). Daigaku nyūshi-no arikata-ni kan suru kentei kaigi teigen (大学入試のあり方に関する検討会議提言) [Recommendations from the study group on the university entrance examination system]. Ministry of Education, Culture, Sports, Science and Technology–Japan. Retrieved from https://www.mext.go.jp/content/20210707-mxt_daigakuc02-000016687_13.pdf

Ministry of Education, Culture, Sports, Science and Technology–Japan. (2021b). Sankō shiryō 3: Daigaku nyūgakusha senbatsu-ni okeru eigo 4 ginō hyōka oyobi kijutsushiki mondai-no jittai chōsa-no kekka (sono 2) (参考資料3:大学入学者選抜における英語４技能評価及び記述式問題の実態調査の結果(その2)) [Reference materials No. 3: Results of the fact-finding surveys on evaluations of English four skills and on written questions used in the university admission selection processes, Part 2]. MEXT. Retrieved from https://www.mext.go.jp/content/20210707-mxt_daigakuc02-000016687_10.pdf

Plonsky, L., Egbert, J., & Laflair, G. T. (2015). Bootstrapping in applied linguistics: Assessing its potential using shared data. *Applied Linguistics*, *36*(5), 591–610. https://doi.org/10.1093/applin/amu001

Puth, M. T., Neuhäuser, M., & Ruxton, G. D. (2015). On the variety of methods for calculating confidence intervals by bootstrapping. *Journal of Animal Ecology*, *84*(4), 892–897. https://doi.org/10.1111/1365-2656.12382

Read, J. (2022). Test review: The International English Language Testing System (IELTS). *Language Testing, 39*(4), 679–694. https://doi.org/10.1177/02655322221086211

Masaya Kanzaki (kanzaki-m@kanda.kuis.ac.jp) has been with Kanda University of International Studies since 2010. His research interests include language testing, vocabulary acquisition, and corpus linguistics.