ets research institute

APRIL
2025

TOEFL® RESEARCH SERIES

Aligning Scores of Language Proficiency Tests

A Score Concordance Study Between IELTS Academic and TOEFL iBT®

Naoki Ikeda, Tony Clark, Spiros Papageorgiou, Lixiong Gu, Renka Ohta, Andrew Blackhurst, & Emma Bruce

TOEFL–RR-105

ETS Research Report
No. RR–25-02

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

Aligning Scores of Language Proficiency Tests: A Score Concordance Study Between IELTS Academic and TOEFL iBT®

Naoki Ikeda,[1] Tony Clark,[2] Spiros Papageorgiou,[3] Lixiong Gu,[3] Renka Ohta,[3]
Andrew Blackhurst,[2] & Emma Bruce[4]

[1]Australian Council for Educational Research, Camberwell, Victoria, Australia
[2]Cambridge University Press & Assessment, Cambridge, United Kingdom
[3]ETS Research Institute, ETS, Princeton, New Jersey, United States
[4]British Council, London, United Kingdom

Abstract

Scores of different language tests intended for similar purposes (e.g., admission to higher education) are used to determine candidates' language proficiency and readiness for a chosen domain. To be fair to all students irrespective of the test they took, score requirements should be comparable. Score concordance tables provide an empirical basis for such comparability when good practice principles are met. We report on a score concordance project whose ambitious goal was to adhere to the good practice principles laid out in Knoch and Fan (2024). The providers of the two most widely-used English language proficiency tests for academic admissions purposes, IELTS Academic and TOEFL iBT®, collaborated to complete this project. Research teams representing both tests recruited 969 test takers who took the tests in a counterbalanced order. The study participants represented the major first language groups of both test-taking populations. Every effort was made to keep the interval between taking the two tests short to minimize any effect of changes in the test-takers' language proficiency level. There were no self-reported score data, a major limitation of existing score concordance studies, as all score reports were verified by the test providers using their official score verification service. Equipercentile equating was conducted by a third party, independent from the test providers. We discuss the challenges in meeting several good practice principles and present the implications for building

trustworthy score concordance tables to help stakeholders make informed decisions about language test acceptance.

Keywords: English language proficiency, IELTS Academic, score concordance, score requirements, test equating, TOEFL iBT®

Corresponding author: S. Papageorgiou, Email: spapageorgiou@ets.org

Acknowledgments

This research was funded by the IELTS Partners: British Council, IDP: IELTS Australia, and Cambridge University Press & Assessment, as well as ETS, owner of the TOEFL® family of tests. The authors are grateful to several colleagues based in local ETS and IDP offices, as well as partner organizations, for organizing data collection sessions in their regions. The authors would also like to thank colleagues who provided feedback on earlier versions of this report, as well as Ayleen Gontz (ETS) and John Savage (Cambridge University Press & Assessment) for copyediting.

Introduction

Users of different language proficiency tests, such as higher education institutions, need to set comparable score requirements. Concordance tables offer a convenient way to compare scores and can help to inform the setting of comparable score requirements (Knoch & Fan, 2024). The study reported here provides IELTS Academic–TOEFL iBT® concordance tables for each test section score (Reading, Listening, Speaking, and Writing) and the Total score (or Overall for IELTS Academic) for which an equipercentile equating procedure was employed. The study participants are described first, followed by an initial data evaluation including descriptive summaries. The concordant results are provided subsequently, where the methodology is detailed, followed by a description of the population invariance study and a discussion of general guidelines for use of the study results. Supplementary information is provided in the appendices.

Overview of the Two Tests

IELTS

IELTS is an international test of English proficiency assessing all four skills: listening, reading, speaking, and writing. The British Council, IDP: IELTS Australia and Cambridge

University Press & Assessment jointly own IELTS. There are two types of IELTS test: IELTS Academic and IELTS General Training. The Listening and Speaking sections are the same for both IELTS tests, but the Reading and Writing sections are different. The Listening, Reading, and Writing sections are completed in one sitting, without breaks. The Speaking section is completed separately, taken within a week before or after the written test. The total test time is 2 hours and 45 minutes, in the sequence of Listening, Reading, and Writing in one sitting plus the Speaking test in a separate sitting as described above.

IELTS is a primarily paper-based test, but it is also offered in a computer-delivered format. Computer-delivered IELTS is the same as the paper-based IELTS in terms of content, structure, question types, marking, test report form, and test timings. However, the test timing for Listening is slightly different. In the paper-based IELTS, test takers need to transfer their answers to an answer sheet, whereas this step is unnecessary in computer-delivered IELTS when test takers can answer directly on computer. The Speaking test remains face to face with a certified IELTS examiner in both in-person and online formats.

Test results are reported on a scale of 0–9 in increments of 0.5 points for the four skills separately, as well as an average score for the whole test.

TOEFL iBT

The TOEFL iBT test, owned by ETS, is the most recent iteration of the TOEFL test. Its purpose is to evaluate the English proficiency of people whose first language is not English. Test scores are primarily used as a measure of the ability of international students to use English in an academic environment. The TOEFL iBT test is administered via computer from a secure, worldwide, internet-based testing network.

The test includes four sections: Reading, Listening, Speaking, and Writing. Each section is scored on a 0–30 scale, resulting in a total possible score of 120. The test takes about 2 hours to complete. Some test tasks require the use of two or more language skills. Test takers speak into a microphone to record their responses to Speaking tasks and type their responses to Writing tasks. The spoken and written responses are digitally recorded and sent to the ETS Online Network for Evaluation (ONE) for scoring.

Test Comparison

For the purposes of this study, scores were compared between IELTS Academic and TOEFL iBT tests. IELTS General Training scores were not the focus of this study. A score

concordance study was conducted by ETS soon after the launch of the TOEFL iBT test in 2005 (ETS, 2010). The results of the present study replace the results of that previous study and include score data following content revisions to the TOEFL iBT test in July 2023 (Davis & Norris, 2023; Gu et al., 2023) and revisions to the IELTS Writing assessment criteria in June 2023 (Clark et al., 2023).

Score concordance tables can be created between any two tests, but scores should not automatically be considered interchangeable unless the two tests are targeting similar test-taker populations, uses, and constructs (Knoch & Fan, 2024). Both tests in this study target the same test-taker groups and are used for the same purpose (largely the evaluation of academic language ability to study in an English-speaking institution of higher education). They also evaluate similar constructs, as evidenced by the inclusion of four sections targeting the four language skills (reading, listening, speaking writing) and the reporting of the same types of subscores.

The two tests also have some noticeable design differences, primarily in the way they evaluate speaking skills. A separate study analyzing the content of the two tests is currently in progress to further establish the extent to which construct comparability between the two tests is sufficient for conducting a score concordance study. Nevertheless, the apparent overlap in constructs, test populations, intended uses, and the existence of a prior concordance study (ETS, 2010) provided a rationale for undertaking a new concordance study.

Sample Description

ETS and the IELTS Partners provided completed score data for 969 test takers who took both tests (TOEFL iBT and IELTS Academic) in test centers (no at-home test administrations were included in this study) between August 2023 and March 2024. All TOEFL iBT and IELTS Academic score reports were confirmed by the test providers using the verification service corresponding to each test; hence, no self-reported scores were included. After carefully examining the data, it was deemed appropriate to exclude some test takers from the analysis. The sample we used comprised data from 937 test takers after we removed those considered to be statistical outliers in the original sample pool (see description below).

A total of 467 (49.8%) test takers took IELTS Academic first, whereas 470 (50.2%) test takers took TOEFL iBT first. The numbers of male and female test takers were 434 (46.3%) and 503 (53.7%), respectively. Their first languages (L1) were diverse, as summarized in Table 1, which also shows a breakdown of the sample used in this report by L1 and by test order. The

SEA language group includes several Southeast Asian languages, and the IND language group includes several languages or dialects in India.

Table 1. Concordance Study Sample by First Language and Test Order

Language	Total *N*	Test order (IELTS–TOEFL)[a]	Test order (TOEFL–IELTS)[a]
Arabic	6	3	3
Chinese	239	118	121
Farsi	1	1	0
French	5	1	4
IND[b]	256	157	99
Japanese	107	56	51
Korean	98	49	49
Other	61	11	50
Portuguese	2	0	2
SEA[c]	140	60	80
Spanish	19	9	10
Turkish	3	2	1
Total	937	467	470

[a]IELTS–TOEFL refers to those who took the IELTS Academic first and TOEFL iBT next. TOEFL–IELTS refers those who took the TOEFL iBT first and IELTS Academic next. [b]IND refers to the language group including several languages or dialects in India. [c]SEA refers to the language group including several Southeast Asian languages.

All test takers took both tests within a 100-day period, except for 14 test takers, for whom the two tests were administered between 101–105 days. Further evaluation of the data showed that 46.2% of the test takers took both tests within 1 month (31 days) and 75.0% of the test takers took both tests within 2 months (62 days). On average the two tests were taken within 38.6 days. The outliers were identified and removed according to the following procedure:

Step 1: All the test takers' scores on both tests were validated by the test providers. Any test takers with missing scores on one or more test sections were removed. This resulted in the data file containing a full set of validated scores for each test taker ($N = 969$).

Step 2: After validation of all test scores and removal of missing scores, regression analyses were run to predict IELTS Academic scores from TOEFL iBT scores and vice versa on each section (measure) and the total, respectively. An absolute value of 3 or larger studentized residual was used as a criterion for flagging outliers on each section and the total scores. Using this criterion, 32 test

takers were flagged on one or more sections or the total and were removed from the data file in Step 1; the resulting cleaned data file contained 937 test takers for concordance analysis.

The participants recruited for this study were a reasonably representative sample of the test-taking populations of both tests, as shown in Table 2. Naturally, the percentage of the various L1 groups in this study cannot be identical to the test-taking populations of both tests. However, Table 2 indicates that the study sample includes major L1 groups for both tests, such as Chinese and Indian speakers. Additionally, this study was conducted in response to a need for comparability of score requirements in the Australian migration context. Therefore, the study sample also included countries and regions that contribute to net overseas migration to Australia, such as India, China, Japan, Korea, and Southeast Asia (Australian Bureau of Statistics, 2023). It should be noted that the tests and their scores in this report do not replace English tests and scores accepted for Australian visa purposes at the time of writing, or indicate acceptance of these tests and their scores by the Department of Home Affairs for Australian visa purposes in the future. Accepted tests and scores can be found on the Department of Home Affairs website.

Table 2. Concordance Study Sample by First Language Compared to the Population of the Two Tests

Language	Concordance study 2023–24 (%)	IELTS Academic 2022 (%)	TOEFL iBT 2022 (%)
Arabic	0.6	5.4	3.5
Chinese	25.5	15.8	35.4
Farsi	0.1	1.4	3.0
French	0.5	0.9	4.1
IND[a]	27.3	39.6	13.2
Japanese	11.4	1.3	6.4
Korean	10.5	0.8	7.8
Other	6.5	23.3	12.5
Portuguese	0.2	0.5	2.0
SEA[b]	14.9	8.2	2.3
Spanish	2.0	1.8	8.6
Turkish	0.3	1.0	1.3
Total	100.0	100.0	100.0

Note. Percentage totals do not always sum due to rounding.
[a]IND: For IELTS, language group includes Pakistani and Bangladeshi speakers of Punjabi, Urdu and Bengali. [b]SEA: language group including several Southeast Asian languages.

Initial Data Evaluation

Table 3 summarizes the following descriptive statistics ($N = 937$) for each test section and Overall/Total scores for each of the two tests: mean, standard deviation, and the observed maximum and minimum scores. IELTS Academic reports each section level and overall scores on a 9-band scale in one-half (0.5) band increments. The study participants' observed scores ranged from 3.5 to 9.0. The reported TOEFL iBT score scale ranges from 0 to 30 for each section and from 0 to 120 for the overall test in 1-point increments. The study participants' section score ranged from 0–30, and the total scores ranged from 20 to 119.

Table 3. Descriptive Summaries of IELTS Academic and TOEFL iBT

Test	Section	Mean	*SD*	Min	Max
IELTS Academic	Reading	6.72	1.25	3.5	9.0
	Listening	6.77	1.24	3.5	9.0
	Speaking	6.20	0.84	4.0	9.0
	Writing	6.15	0.59	4.5	8.0
	Overall	6.52	0.86	4.5	8.5
TOEFL iBT	Reading	19.25	7.38	0	30
	Listening	19.22	7.07	1	30
	Speaking	20.02	4.26	6	30
	Writing	19.33	5.02	4	30
	Total	77.82	20.77	20	119

Note: $N = 937$. Min = minimum; Max = maximum

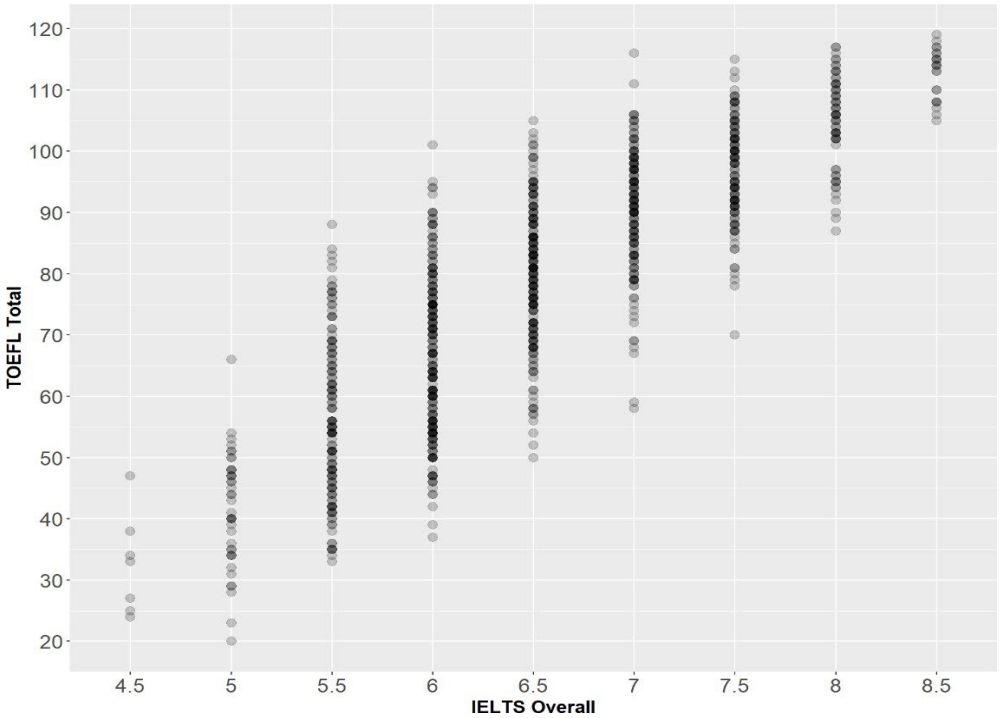
The number of test-takers for each score level and corresponding percentages for each section and Overall/Total for IELTS Academic and TOEFL iBT are provided in Appendix A. The score correlations between the two tests were .76 (Reading), .70 (Listening), .69 (Speaking), .68 (Writing), and .85 (Overall/Total). Figure 1 shows the scatter plot (the bivariate distribution) for the Overall/Total scores of IELTS Academic and TOEFL iBT.

Given the observed score ranges, the spread of the test takers' scores, the mean scores for the two tests as well as the diverse language background of the test takers, the current sample was considered a reasonable representation of the test-taker population of interest. The correlations between the two tests (in each section and in the Overall/ Total) were moderate to strong and stronger than those reported in the previous study conducted by ETS (ETS, 2010).

Further, a possible effect of test order (whether IELTS Academic or TOEFL iBT were taken first) was explored descriptively and by subsequent statistical analyses. Across the sections and Overall/Total scores of the two tests, the group of test-takers who took TOEFL iBT first

outperformed the other group (i.e., those who took IELTS Academic first). The mean difference between the two groups ranged from 1.26 to 3.22 (out of 30 points) for the TOEFL iBT four section scores; from 0.09 to 0.30 (out of 9 points) for the IELTS Academic four section scores; 9.25 (out of 120 points) for the TOEFL iBT Total score; and 0.19 (out of 9 points) for IELTS Academic Overall score.

Figure 1. The Observed Distribution of the Two Tests



Subsequently, a multivariate analysis of variance (MANOVA) was performed on the IELTS Academic and TOEFL iBT scores. For each analysis, test order was defined as the independent variable and the four section scores were defined as a set of dependent variables. Additionally, an analysis of variance (ANOVA) was performed for IELTS Academic Overall and TOEFL iBT Total scores respectively, where test order was defined as the independent variable and the test score (Overall/Total) was defined as the dependent variable. In all analyses, test order was found to significantly account for the test scores. The effect of test order on the two tests was, however, found to be small, with the effect size (η_p^2 or η^2) less than .06. Thus, the current study did not find strong evidence that test order substantially advantaged or disadvantaged the sample of test takers on the respective tests.

Score Concordance Results

Equipercentile linking procedures were performed for the four sections and Overall/ Total score. The equipercentile linking method has been used in other studies on linking between internationally recognized English language tests, such as Clesham and Hughes (2020) for PTE–IELTS, ETS (2010) for TOEFL–IELTS, and Saville et al. (2021) for IELTS–PTE. Given the sample size, which is relatively small, it was deemed appropriate to perform a presmoothing, the use of which is well justified by Saville et al. (2021). As practiced in Clesham and Hughes, log-linear presmoothing was applied under a single group design (i.e., smoothing was done for bivariate data, instead of for separate univariate data). Model fit of different log-linear models was evaluated using Akaike Information Criterion (AIC) (Akaike, 1981). These statistical operations were performed using an R package, *equate* (Albano, 2016).

The following figures show the smoothed (solid line) and unsmoothed (dash line) distributions. Figures 2 to 4 show the distributions for the four section scores and the Overall score for IELTS Academic. Figures 5 to 7 show the distributions for the four section scores and the Total score for TOEFL iBT.

Figure 2. Distribution: Reading and Listening (IELTS Academic)

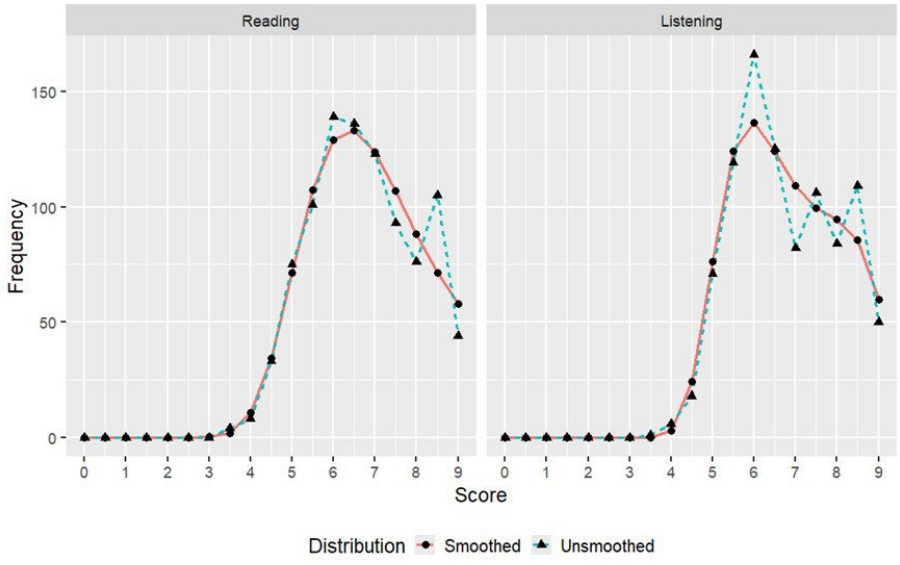


Figure 3. Distribution: Speaking and Writing (IELTS Academic)

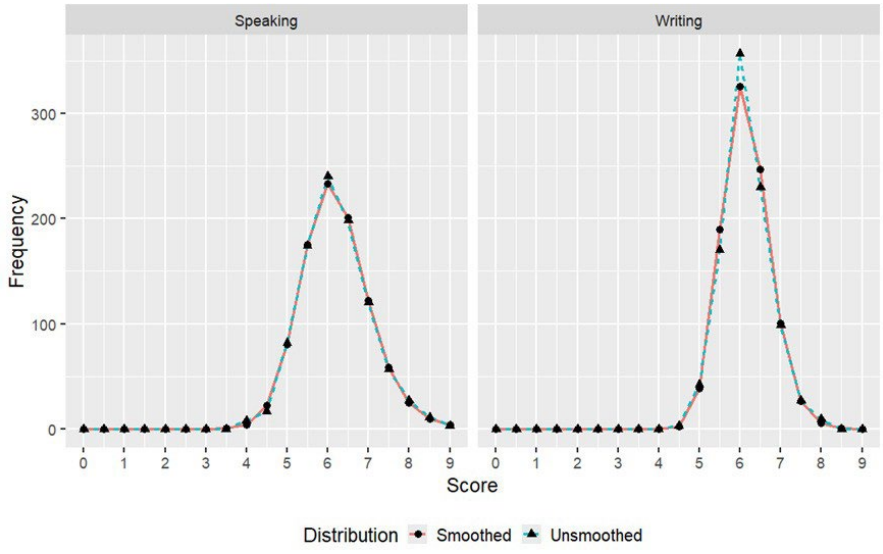


Figure 4. Distribution: Overall (IELTS Academic)

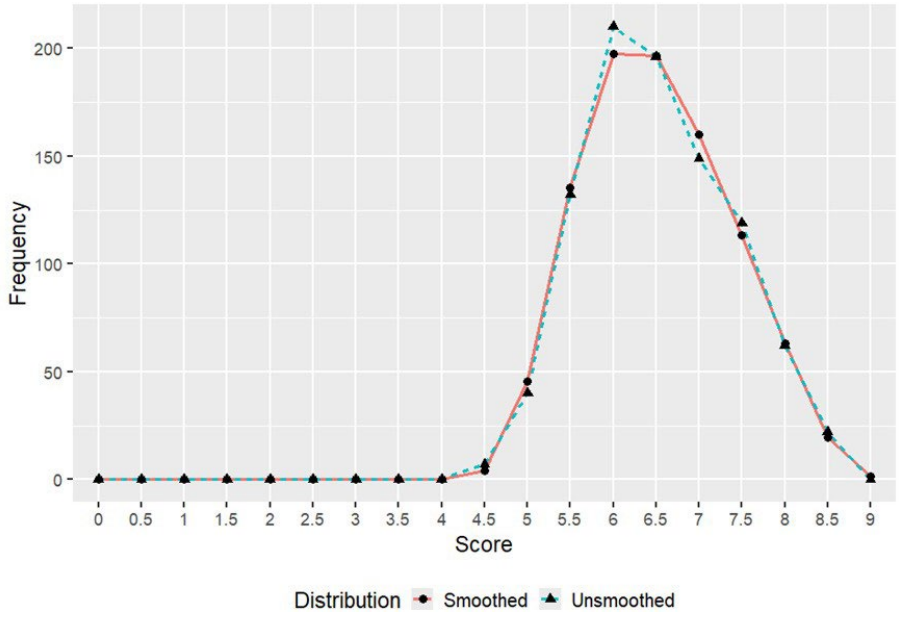


Figure 5. Distribution: Reading and Listening (TOEFL iBT)

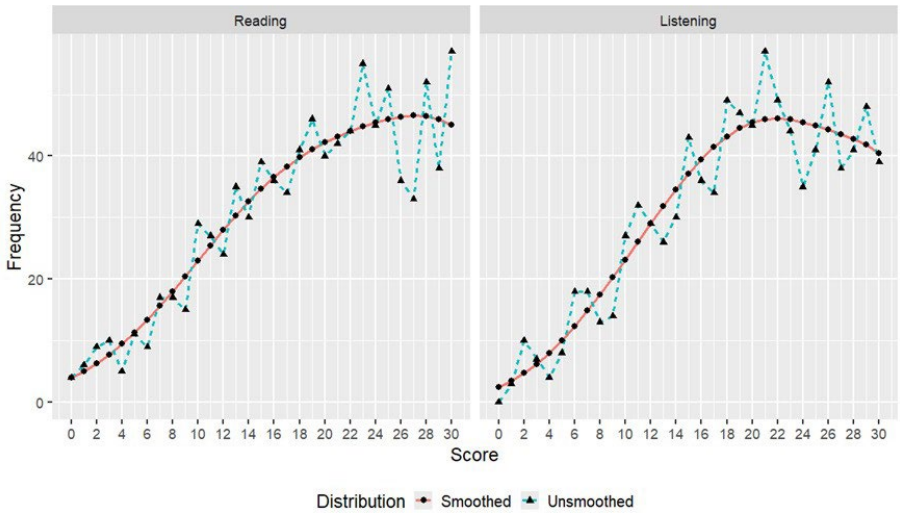


Figure 6. Distribution: Speaking and Writing (TOEFL iBT)

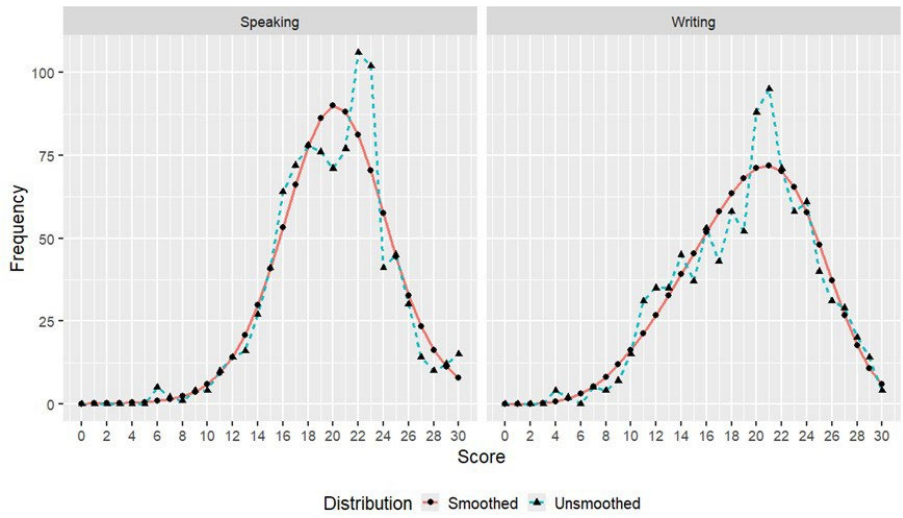
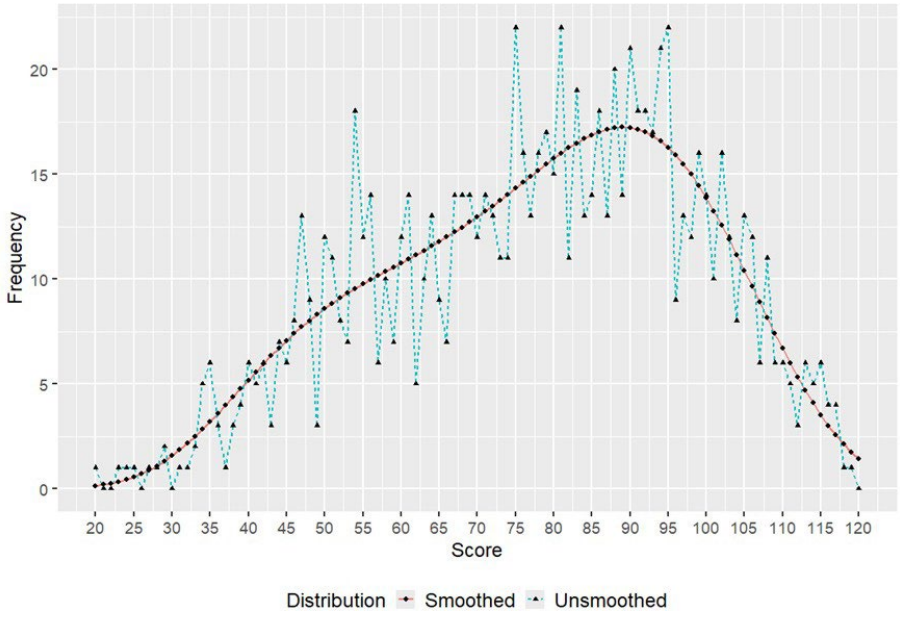


Figure 7. Distribution: Total (TOEFL iBT)



The concordant results are provided in Tables 4 to 8 by IELTS Academic half bands (Band 4 to 9) and the corresponding TOEFL iBT score ranges. Table 4 presents the concordant results for Reading on the two tests. The first column shows the IELTS Academic band and the second column shows the number of test takers classified at a particular band. The range of the corresponding TOEFL iBT scores based on the equipercentile method is provided in the third column. The last column shows the standard error, defined as the standard deviation of all TOEFL iBT scores divided by the square root of the sample size at the corresponding IELTS Academic band.

Tables 5 to 8 present the results for Listening, Speaking, Writing, and Overall/Total respectively. The tables can be interpreted in the same way as described for Table 4. Given the motivation for conducting this study, as mentioned earlier, Appendix B included the concordant results organized by proficiency levels used by the Australian Government (Australian Government, Department of Home Affairs, 2024a, 2024b). As mentioned earlier, the results in Appendix B do not replace English tests and scores accepted for Australian visa purposes at the time of writing or indicate acceptance of these tests and their scores by the Department of Home Affairs for Australian visa purposes in the future.

Table 4. Concordance Table for Reading (at IELTS Academic 0.5 Levels)

IELTS Academic score	*n*	TOEFL iBT score range	*SE*
9.0	44	30	0.41
8.5	105	28–29	0.32
8.0	76	27	0.43
7.5	93	25–26	0.48
7.0	123	22–24	0.40
6.5	136	19–21	0.42
6.0	139	16–18	0.47
5.5	101	12–15	0.57
5.0	75	8–11	0.62
4.5	33	4–7	0.87
4.0	8	1–3	1.71

Table 5. Concordance Table for Listening (at IELTS Academic 0.5 Levels)

IELTS Academic score	*n*	TOEFL iBT score range	*SE*
9.0	50	30	0.31
8.5	109	28–29	0.36
8.0	84	26–27	0.45
7.5	106	24–25	0.52
7.0	82	22–23	0.56
6.5	125	19–21	0.48
6.0	166	16–18	0.47
5.5	119	12–15	0.50
5.0	71	8–11	0.65
4.5	18	3–7	0.82
4.0	6	0–2	1.69

Table 6. Concordance Table for Speaking (at IELTS Academic 0.5 Levels)

IELTS Academic score	*n*	TOEFL iBT score range	*SE*
9.0	3	30	0.67
8.5	11	29	0.68
8.0	27	28	0.52
7.5	57	26–27	0.30
7.0	120	24–25	0.29
6.5	198	22–23	0.21
6.0	240	19–21	0.21
5.5	174	17–18	0.24
5.0	82	14–16	0.37
4.5	17	11–13	0.99
4.0	8	7–10	1.60

Table 7. Concordance Table for Writing (at IELTS Academic 0.5 Levels)

IELTS Academic score	*n*	TOEFL iBT score range	*SE*
9.0	–	30	–
8.5	–	30	–
8.0	9	30	0.55
7.5	27	28–29	0.46
7.0	99	26–27	0.29
6.5	230	23–25	0.23
6.0	357	19–22	0.21
5.5	170	14–18	0.31
5.0	42	9–13	0.50
4.5	3	4–8	1.20
4.0	–	1–3	–

Table 8. Concordance Table for Overall/Total (at IELTS Academic 0.5 Levels)

IELTS Academic score	*n*	TOEFL iBT score range	*SE*
9.0	–	120	–
8.5	22	115–119	0.89
8.0	62	108–114	0.95
7.5	119	100–107	0.75
7.0	149	91–99	0.80
6.5	196	81–90	0.79
6.0	210	67–80	0.89
5.5	132	51–66	1.13
5.0	40	37–50	1.48
4.5	7	26–36	3.09
4.0	–	14–25	–

Population Invariance

Population invariance was explored to evaluate the adequacy of the linking. The whole sample was split into two groups: male test takers (n = 434) and female test-takers (n = 503). Although other variables of test takers' backgrounds were available (for example L1), the resulting groups were either unbalanced and/or of small sizes. Therefore, the analysis was conducted based on the above two groups. The equated TOEFL iBT scores (unrounded) across the IELTS Academic band levels were estimated separately for each of the male and female groups and compared to the corresponding estimates from the combined (whole) group. Log-linear presmoothing was performed, where the best model was selected for the male and female

data respectively by referring to the AIC values. Figures 8 and 9 visually present the differences in equivalents (male, female, and the whole sample). Although differences are most noticeable with IELTS Academic Bands 3.0 and 4.0 for Speaking and Overall/Total, as well as Bands 7.5 to 8.5 for Overall/Total, the two subgroups and the whole sample produced reasonably similar tendencies (in terms of increase of TOEFL iBT scores across IELTS Academic bands) across the four section and Overall/Total scores. Also note most score-based decisions are not typically based on IELTS Academic Band 4.0 and below.

Figure 8. Unrounded TOEFL iBT Equipercentile Equivalents (Four Sections)

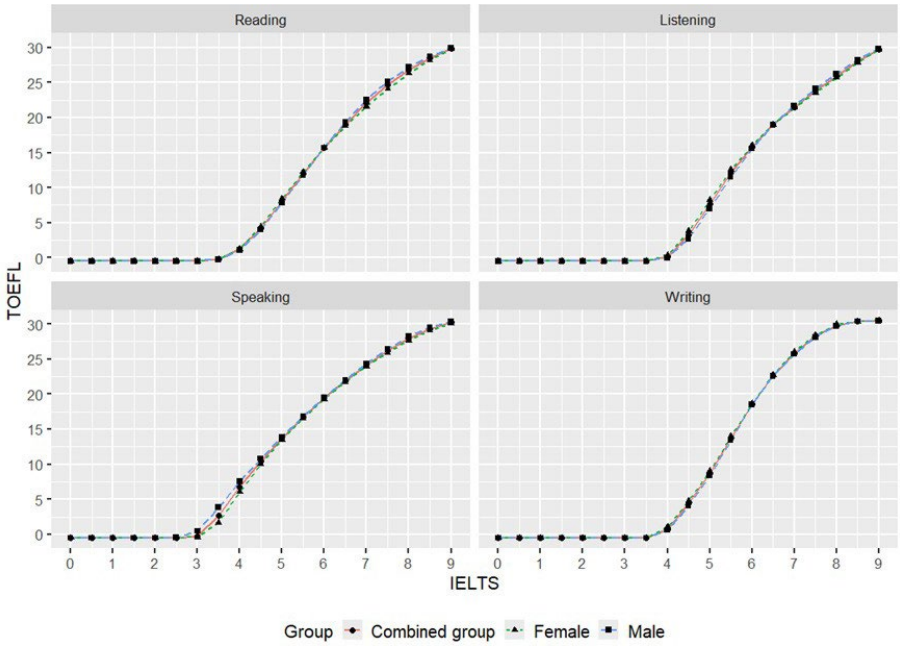
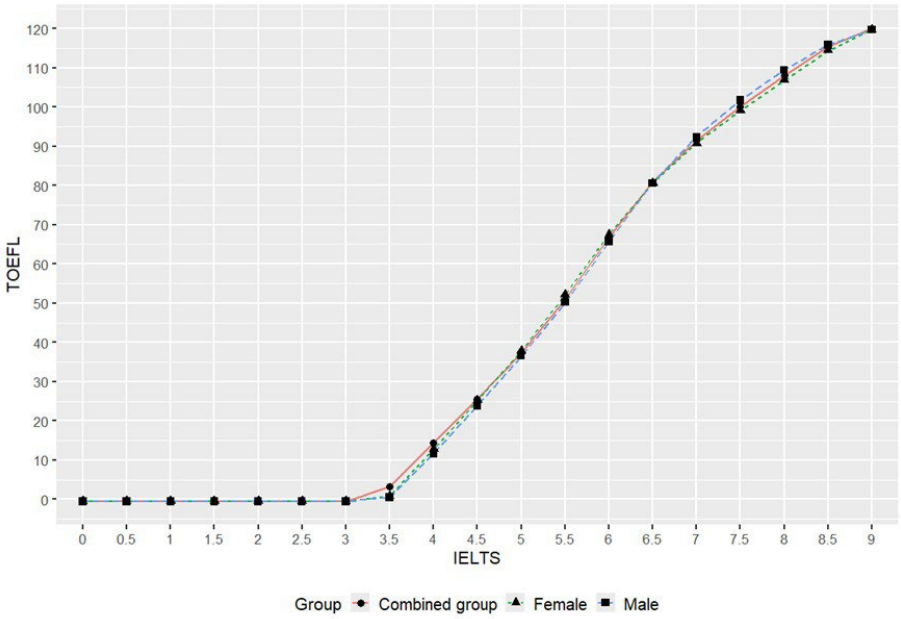


Figure 9. Unrounded TOEFL iBT Equipercentile Equivalents (Overall/Total)



Discussion and Conclusions

This report presents the score concordance results based on the scores of 937 test takers who took both the IELTS Academic and the TOEFL iBT tests in a counterbalanced design. The study participants represented major L1 groups of both tests, as well as countries contributing to net overseas migration to Australia (as mentioned elsewhere, the score concordance results do not replace test acceptance and score requirements for Australian visa purposes at the time or writing or in the future). The robustness of the score concordance results is supported by the methodology of the study (e.g., L1 group representation, counterbalancing, score verification for both tests), as well as information about standard at critical proficiency levels for important, score-based decisions, and a population invariance study.

Given the project timeline and the simultaneous data collection across several countries, collection of detailed familiarity data was not possible. In a subset of the participants (n =149 [15.9%]) who were directly recruited by the ETS research team via email, 108 responded that they prepared for TOEFL iBT and 107 that they prepared for IELTS Academic. Although test familiarity data were not collected for most of the participants, the effect of test familiarity, if any, should cancel out at the overall test-taker group level because of the carefully counterbalanced design (see information about test order in Table 1).

Score concordance tables are useful instruments for score users; for example, institutions who use certain test scores for decisions about test takers and who need to set comparable score requirements across different tests. Score users are advised that score comparisons across tests, although based on empirical research, are estimates only and should be treated with caution for the reasons outlined on the following pages.

- Tests differ, sometimes significantly, in the ways information about English language ability is elicited and assessed. Score comparisons are only meaningful to the extent that the tests are measuring the same ability or skill. As mentioned earlier, both tests in this study target the same test-taker groups and are used for the same purpose. They also evaluate similar constructs, as evidenced by the inclusion of four sections targeting the four language skills (reading, listening, speaking, writing) and the reporting of the same types of subscores. The two tests also have some noticeable design differences, primarily in the way they evaluate speaking skills. For this reason, score users should seek evidence of construct comparability. A separate study analyzing the content of the two tests is currently in progress to confirm that construct comparability is sufficient for the purposes of conducting a score concordance study.
- Tests often differ in the length of the reporting scales used (for example, one test may report on a 6-point scale and another on a 100-point scale). As a result, a one-to-one mapping of scores from one test to another is rarely possible. Such differences in the score reporting mechanism when interpreting results from score concordance studies constitute an important consideration. The TOEFL iBT test uses a 0–120 total score scale and a 30-point score scale for the four test sections. The total score is the sum of the four section scores. IELTS Academic uses a 9-band reporting system with half bands reported for the overall score, which is the average of the section scores. Because of these different score reporting mechanisms, there is no one-to-one mapping of scores from one test to the other, as multiple TOEFL iBT scores may be equivalent to the same IELTS Academic band score. Score users need to consider which of the TOEFL iBT scores corresponding to the same IELTS Academic band is more suitable in their context.
- The choice of concordance study methodology may produce variations in results. Every effort was made when designing and executing this study to adhere to good practice principles, such as those outlined in Knoch and Fan (2024). In addition, the analysis of

the score data was conducted by the first author, who was independent from the providers of the two tests. Involving a third party in a score concordance study is advisable to increase confidence in the resulting concordance tables.

- The populations of test-takers may differ (e.g., with respect to ages, nationalities, language backgrounds of test-takers) from the population used in the research that generated the score equivalences. As discussed earlier in this report, the score concordance population included groups that are also representative of the overall test-taking population of each test.
- The sample sizes used for comparing scores from different tests are generally small for individual levels/ranges, especially at extreme ends of the score scale. In the case of this study, we analyzed score data from both tests taken by almost 1,000 test-takers, as suggested in the literature (Knoch & Fan, 2024). It should be noted that all scores were verified by the test providers, supporting the quality of the collected data.
- Score concordance results are generally more robust for proficiency levels with larger numbers of test-takers. In this study, larger test-taker numbers were observed with IELTS Academic Overall Bands 5.5–7.5 (100 test-takers or more per half band), which reflects the operational test-taking population, and the key decision points when using test scores.
- Large standard errors show that score equivalences are particularly imprecise at certain points on the ability scale. All tables in this report include the standard error, defined as the standard deviation of all TOEFL iBT scores divided by the square root of the sample size at the corresponding IELTS Academic half band.
- Because the score comparisons presented in the score comparison tables are indicative only, score users are advised not to rely solely on published score equivalences in making their decisions. They should weigh evidence from additional sources where feasible. The providers of both IELTS Academic and TOEFL iBT offer additional materials to help score users set useful and relevant score requirements (ETS, 2020; IELTS, 2014). Such materials advocate the use of various types of information, beyond score concordance tables.

Appendix A

The number of test-takers for each score level and the corresponding percentages for each section and Overall/Total for IELTS Academic and TOEFL iBT.

Table A1. IELTS Academic Score Distribution (Four Sections)

Score	Reading (*n*)	Listening (*n*)	Speaking (*n*)	Writing (*n*)	Reading (%)	Listening (%)	Speaking (%)	Writing (%)
9.0	44	50	3	0	4.7	5.3	0.3	0.0
8.5	105	109	11	0	11.2	11.6	1.2	0.0
8.0	76	84	27	9	8.1	9.0	2.9	1.0
7.5	93	106	57	27	9.9	11.3	6.1	2.9
7.0	123	82	120	99	13.1	8.8	12.8	10.6
6.5	136	125	198	230	14.5	13.3	21.1	24.5
6.0	139	166	240	357	14.8	17.7	25.6	38.1
5.5	101	119	174	170	10.8	12.7	18.6	18.1
5.0	75	71	82	42	8.0	7.6	8.8	4.5
4.5	33	18	17	3	3.5	1.9	1.8	0.3
4.0	8	6	8	0	0.9	0.6	0.9	0.0
3.5	4	1	0	0	0.4	0.1	0.0	0.0

Note. Across the sections, no test taker's score fell within the range of 0–3.0.

Table A2. IELTS Academic Score Distribution (Overall)

Score	*n*	%
8.5	22	2.3
8.0	62	6.6
7.5	119	12.7
7.0	149	15.9
6.5	196	20.9
6.0	210	22.4
5.5	132	14.1
5.0	40	4.3
4.5	7	0.7

Note. No test taker's score fell within the range of 0–4.0.

Table A3. TOEFL iBT Score Distribution (Four Sections)

Score	Reading (*n*)	Listening (*n*)	Speaking (*n*)	Writing (*n*)	Reading (%)	Listening (%)	Speaking (%)	Writing (%)
30	57	39	15	4	6.1	4.2	1.6	0.4
29	38	48	12	14	4.1	5.1	1.3	1.5
28	52	41	10	20	5.5	4.4	1.1	2.1
27	33	38	14	29	3.5	4.1	1.5	3.1
26	36	52	30	31	3.8	5.5	3.2	3.3
25	51	41	45	40	5.4	4.4	4.8	4.3
24	45	35	41	61	4.8	3.7	4.4	6.5
23	55	44	102	58	5.9	4.7	10.9	6.2
22	44	49	106	71	4.7	5.2	11.3	7.6
21	42	57	77	95	4.5	6.1	8.2	10.1
20	40	45	71	88	4.3	4.8	7.6	9.4
19	46	47	76	52	4.9	5.0	8.1	5.5
18	41	49	78	58	4.4	5.2	8.3	6.2
17	34	34	72	43	3.6	3.6	7.7	4.6
16	36	36	64	53	3.8	3.8	6.8	5.7
15	39	43	41	37	4.2	4.6	4.4	3.9
14	30	30	27	45	3.2	3.2	2.9	4.8
13	35	26	16	35	3.7	2.8	1.7	3.7
12	24	29	14	35	2.6	3.1	1.5	3.7
11	27	32	10	31	2.9	3.4	1.1	3.3
10	29	27	4	15	3.1	2.9	0.4	1.6
9	15	14	4	7	1.6	1.5	0.4	0.7
8	17	13	1	4	1.8	1.4	0.1	0.4
7	17	18	2	5	1.8	1.9	0.2	0.5
6	9	18	5	0	1.0	1.9	0.5	0.0
5	11	8	0	2	1.2	0.9	0.0	0.2
4	5	4	0	4	0.5	0.4	0.0	0.4
3	10	7	0	0	1.1	0.7	0.0	0.0
2	9	10	0	0	1.0	1.1	0.0	0.0
1	6	3	0	0	0.6	0.3	0.0	0.0
0	4	0	0	0	0.4	0.0	0.0	0.0

Table A4. TOEFL iBT Score Distribution (Total)

Score	*n*	%	Score	*n*	%	Score	*n*	%	Score	*n*	%
120	0	0.0	90	21	2.2	60	12	1.3	30	0	0.0
119	1	0.1	89	14	1.5	59	7	0.7	29	2	0.2
118	1	0.1	88	20	2.1	58	10	1.1	28	1	0.1
117	4	0.4	87	13	1.4	57	6	0.6	27	1	0.1
116	4	0.4	86	18	1.9	56	14	1.5	26	0	0.0
115	6	0.6	85	14	1.5	55	12	1.3	25	1	0.1
114	5	0.5	84	13	1.4	54	18	1.9	24	1	0.1
113	6	0.6	83	19	2.0	53	7	0.7	23	1	0.1
112	3	0.3	82	11	1.2	52	8	0.9	22	0	0.0
111	5	0.5	81	22	2.3	51	11	1.2	21	0	0.0
110	6	0.6	80	15	1.6	50	12	1.3	20	1	0.1
109	6	0.6	79	17	1.8	49	3	0.3	19	0	0.0
108	11	1.2	78	16	1.7	48	9	1.0	18	0	0.0
107	6	0.6	77	13	1.4	47	13	1.4	17	0	0.0
106	12	1.3	76	16	1.7	46	8	0.9	16	0	0.0
105	13	1.4	75	22	2.3	45	6	0.6	15	0	0.0
104	8	0.9	74	11	1.2	44	7	0.7	14	0	0.0
103	12	1.3	73	11	1.2	43	3	0.3	13	0	0.0
102	16	1.7	72	13	1.4	42	6	0.6	12	0	0.0
101	10	1.1	71	14	1.5	41	5	0.5	11	0	0.0
100	14	1.5	70	12	1.3	40	6	0.6	10	0	0.0
99	16	1.7	69	14	1.5	39	4	0.4	9	0	0.0
98	12	1.3	68	14	1.5	38	3	0.3	8	0	0.0
97	13	1.4	67	14	1.5	37	1	0.1	7	0	0.0
96	9	1.0	66	7	0.7	36	3	0.3	6	0	0.0
95	22	2.3	65	9	1.0	35	6	0.6	5	0	0.0
94	21	2.2	64	13	1.4	34	5	0.5	4	0	0.0
93	17	1.8	63	10	1.1	33	2	0.2	3	0	0.0
92	18	1.9	62	5	0.5	32	1	0.1	2	0	0.0
91	18	1.9	61	14	1.5	31	1	0.1	0–1	0	0.0

Appendix B

The concordant results provided in Tables B1–B5 are organized by proficiency levels used by the Australian Government (Australian Government, Department of Home Affairs, 2024a, 2024b), given the motivation for conducting this study, as mentioned earlier. Table B1 presents the concordant results for Reading on the two tests. The first column shows the proficiency levels, and the second column shows the number of study participants classified at a particular level based on the IELTS Academic band they achieved. The IELTS Academic bands needed to be classified at each level are shown in the third column. It should be noted that the number of test takers in the second column refers to the proficiency level used by the Australian Government, not necessarily the minimum IELTS Academic band required for each level. For example, 275 test takers achieved either an IELTS Academic band of 6 or 6.5, classifying them as Competent level (the next level up, Proficient, requires a minimum IELTS Academic band of 7). The range of the corresponding TOEFL iBT scores based on the equipercentile method is provided in the fourth column. The last column shows the standard error, defined as the standard deviation of all TOEFL iBT scores divided by the square root of the sample size at the corresponding proficiency levels.

Tables B2 to B5 present the results for Listening, Speaking, Writing, and Overall/Total respectively. The tables can be interpreted in the same way as described for Table B1. It should be noted that visa-specific scores are subject to change. The tests and their scores in this report do not replace English tests and scores accepted for Australian visa purposes at the time of writing or indicate acceptance of these tests and their scores by the Department of Home Affairs for Australian visa purposes in the future. Visa applicants need to check the Department of Home Affairs (DHA) current English language requirements for the visa subclass they wish to apply for.

Table B1. Concordance Table for Reading

Proficiency level	*n*	IELTS Academic score	TOEFL iBT score range	*SE*
Below Functional	8	4	1–3	1.71
Functional	33	4.5	4–7	0.87
Vocational	75	5	8–11	0.62
Visa-specific scores	101	5.5	12–15	0.57
Competent	275	6	16–21	0.33
Proficient	216	7	22–26	0.31
Superior	225	8	27–30	0.23

Note. The names of these proficiency levels, with the exception of Below Functional, were set by the Department of Home Affairs.

Table B2. Concordance Table for Listening

Proficiency level	*n*	IELTS Academic score	TOEFL iBT score range	*SE*
Below Functional	6	4	0–2	1.69
Functional	18	4.5	3–7	0.82
Vocational	71	5	8–11	0.65
Visa-specific scores	119	5.5	12–15	0.50
Competent	291	6	16–21	0.35
Proficient	188	7	22–25	0.38
Superior	243	8	26–30	0.25

Note. The names of these proficiency levels, with the exception of Below Functional, were set by the Department of Home Affairs.

Table B3. Concordance Table for Speaking

Proficiency level	*n*	IELTS Academic score	TOEFL iBT score range	*SE*
Below Functional	8	4	7–10	1.60
Functional	17	4.5	11–13	0.99
Vocational	82	5	14–16	0.37
Visa-specific scores	174	5.5	17–18	0.24
Competent	438	6	19–23	0.15
Proficient	177	7	24–27	0.22
Superior	41	8	28–30	0.42

Note. The names of these proficiency levels, with the exception of Below Functional, were set by the Department of Home Affairs.

Table B4. Concordance Table for Writing

Proficiency level	*n*	IELTS Academic score	TOEFL iBT score range	*SE*
Below Functional	-	4	1–3	-
Functional	3	4.5	4–8	1.20
Vocational	42	5	9–13	0.50
Visa-specific scores	170	5.5	14–18	0.31
Competent	587	6	19–25	0.17
Proficient	126	7	26–29	0.25
Superior	9	8	30	0.55

Note. The names of these proficiency levels, with the exception of Below Functional, were set by the Department of Home Affairs.

Table B5. Concordance Table for Overall/Total

Proficiency level	*n*	IELTS Academic score	TOEFL iBT score range	*SE*
Below Functional	-	4	14–25	-
Functional	7	4.5	26–36	3.09
Vocational	40	5	37–50	1.48
Visa-specific scores	132	5.5	51–66	1.13
Competent	406	6	67–90	0.68
Proficient	268	7	91–107	0.59
Superior	84	8	108–120	0.82

Note. The names of these proficiency levels, with the exception of Below Functional, were set by the Department of Home Affairs.

References

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, *16*(1), 3–14. https://doi.org/10.1016/0304-4076(81)90071-3

Albano, A. D. (2016). equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, *74*(8), 1–36. https://doi.org/10.18637/jss.v074.i08

Australian Bureau of Statistics. (2023, December 15). *Overseas migration. Statistics on Australia's international migration, by state and territory, country of birth, visa, age and sex: 2022–23 financial year.* https://www.abs.gov.au/statistics/people/population/overseas- migration/2022-23-financial-year

Australian Government, Department of Home Affairs. (2024a). *English language visa requirements*. https://immi.homeaffairs.gov.au/help-support/meeting-our-requirements/english-language

Australian Government, Department of Home Affairs. (2024b). *New English language requirements*. https://www.homeaffairs.gov.au/news-media/archive/article?itemId=1188

Clark, T., Tasviri, R., & Bruce, E. (2023). *IELTS Writing scales review and summary overview*. IELTS. https://ielts.org/researchers/our-research/research-reports/ielts-writing-scales-review-and-update-summary-overview

Clesham, R., & Hughes, S. R. (2020). *2020 concordance report: PTE Academic and IELTS Academic*. Pearson. https://www.pearsonpte.com/ctf-assets/yqwtwibiobs4/1hXHbkTLYCJly7JryACWjK/5a20dbe26d8ca2c36a3b0dd5a32868d7/2021_PTEA_2020_ PTE_IELTS_Concordance_White_Paper.pdf

Davis, L., & Norris, J. M. (2023). *A comparison of two TOEFL® writing tasks* (Research Memorandum No. RM-23-06). ETS. www.ets.org/Media/Research/pdf/RM-23-06.pdf

ETS. (2010). *Linking TOEFL iBT® scores to IELTS® scores: A research report*. ETS. https://www.ets.org/pdfs/toefl/linking-toefl-ibt-scores-to-ielts-scores.pdf

ETS. (2020). *Guidelines for setting useful score requirements for the TOEFL iBT® test. Volume 9.* ETS. https://www.ets.org/pdfs/toefl/toefl-ibt-insight-s1v9.pdf

Gu, L., Li, S., Li, T., & Norris, J. M. (2023). *Maintaining score quality on the enhanced TOEFL iBT® test* (Research Memorandum No. RM-23-05). ETS. https://www.ets.org/Media/Research/pdf/RM-23-05.pdf

IELTS. (2014). *IELTS scores guide*. https://s3.eu-west-2.amazonaws.com/ielts-web-static/production/Guides/ielts-scores-guide.pdf

Knoch, U., & Fan, J. (2024). Test score comparison tables: How well are they serving test users? *Language Testing*, *41*(3), 681–693. https://doi.org/10.1177/02655322241239348

Saville, N., O'Sullivan, B., & Clark. T. (Eds.). (2021). Investigating the relationship between IELTS and PTE-Academic. *IELTS Partnership Research Papers: Studies in Test Comparability Series No. 2*. British Council/Cambridge Assessment English/IDP: IELTS Australia. https://ielts.org/researchers/our-research/research-reports/investigating-the-relationship-between-pte-academic-and-ielts-academic

Suggested Citation

Ikeda, N., Clark, T., Papageorgiou, S., Gu., L., Ohta, R., Blackhurst, A., & Bruce, E. (2025). *Aligning scores of language proficiency tests: A score concordance study between IELTS Academic and TOEFL iBT*® (TOEFL Research Report No. RR-105). ETS. https://www.ets.org/Media/Research/pdf/RR-25-02

This report is also available as an IELTS publication: Ikeda, N., Clark, T., Papageorgiou, S., Gu, L., Ohta, R., Blackhurst, A., & Bruce, E. (2025). *Aligning scores of language proficiency tests: A score concordance study between IELTS Academic and TOEFL iBT* (IELTS Partnership Research Papers: Studies in Test Comparability Series, No. 1/25). British Council, IDP: IELTS Australia, and Cambridge University Press & Assessment. https://ielts.org/researchers/our-research/research-reports

Funding

This research was funded by the IELTS Partners: British Council, IDP: IELTS Australia, and Cambridge University Press & Assessment, as well as Educational Testing Service (ETS), owner of the TOEFL test.

Action Editor: Jonathan Schmidgall

Reviewers: Larry Davis and Vanessa Manna