

A Concordance Study between the DET and IELTS Academic SWRL Subscores



Duolingo Research Report DRR-24-05
October 24, 2025 (10 pages)
<https://englishtest.duolingo.com/research>

Steven W. Nydick and J.R. Lockwood

Abstract

In 2022, we published an updated concordance between the DET overall score and each of the TOEFL iBT and IELTS Academic overall scores. As of July 1, 2024, DET score reports now include SWRL (i.e., speaking, writing, reading, and listening) component scores in addition to the overall score and integrated skills (i.e., literacy, comprehension, conversation, production) scores. This report updates the previous study to construct a concordance between each of the SWRL components of the DET and IELTS Academic tests. We include sensitivity analyses to account for differences between feasible methods as well as estimated bootstrapped standard errors for the chosen method.

Last Update: October 2025

Keywords

Duolingo English Test, achievement tests, admissions testing, concordance, English proficiency testing, equating, higher education, test validity, SWRL

Contents

1	Introduction	2
2	Methods	2
3	Results	3
3.1	Descriptive Statistics	3
3.2	Concordance Comparisons	4
3.3	Speaking	4
3.4	Writing	5
3.5	Reading	5
3.6	Listening	5
3.7	Overall Score	7
4	Conclusion	7
5	References	9
6	Appendices	10
6.1	Bootstrapping	10

*Duolingo, Inc.

Corresponding author:

Steven W. Nydick

Duolingo, Inc. 5900 Penn Ave, Pittsburgh, PA 15206, USA

Email: englishtest-research@duolingo.com

1 Introduction

Duolingo regularly compares Duolingo English Test (DET) test performance against alternative English language proficiency (ELP) assessments, such as the TOEFL iBT (TOEFL, [ETS, 2012](#)) and IELTS Academic (IELTS, [O’Sullivan, 2018](#)) assessments. Concordance tables comparing different assessments facilitate institutional decisions across test takers who choose to take different tests to demonstrate ELP. Relatively recent concordance studies across ELP tests include the concordance between TOEFL and IELTS assessments ([Ikeda et al., 2025](#)), the concordance between IELTS and PTE assessments ([Yu, 2021](#)), and the concordance between the DET and TOEFL and IELTS assessments ([Cardwell et al., 2023](#)).

[Cardwell et al. \(2023\)](#) included justification for a concordance between the DET and TOEFL/IELTS assessments but only compared overall scores. Institutions often make decisions based on the overall score and one or more subscores. However, the DET has historically only reported integrated subscores (literacy, comprehension, conversation, production) ([LaFlair, 2020](#)) rather than the individual speaking, writing, reading, and listening (SWRL) subscores that other ELP tests have reported. The historical difference in types of subscores between the DET and other ELP tests prevented the previous DET concordance study from including subscore concordance tables.

Some changes to the structure and DET test taking population have occurred since the most recent concordance that have implications for subscore concordance. First, the DET began reporting individual subscores on July 1, 2024, including speaking, writing, reading, and listening (SWRL) scores. Moreover, we now report several composite scores: integrated subscores calculated from an average of pairs of individual subscores; and an overall score constructed from an average of the reported individual subscores (see [Nydick & Lockwood, 2024](#) for more information about the methods of administration and scoring for the DET version corresponding to the concordance). Moreover, as more institutions accept the DET as a demonstration of ELP, fewer test takers report taking an alternative ELP assessment. Those test takers who report multiple assessments show evidence of being different from the general population of DET test takers. This change in test taker behavior requires targeted data collection, assumptions, and statistical adjustments to ensure that any comparisons reflect general test-taker performance.

This report explains the process behind constructing the concordance tables between the DET and IELTS Academic assessments SWRL scores to assist institutions in making comparable decisions across different tests where subscores are used as part of those decisions. We also validate the findings of the most recent concordance study conducted between the DET and IELTS Academic assessments (as reported in [Cardwell et al., 2023](#)).

2 Methods

This concordance followed the design outlined in [Cardwell et al. \(2023\)](#) and most of the good practice principles from [Knoch & Fan \(2024\)](#). We contacted DET test takers who took the DET between April 26, 2024 and September 13, 2024 to ask for official TOEFL iBT or IELTS Academic score reports that we could include in the concordance study. Unlike [Cardwell et al. \(2023\)](#), we included only official score reports in the final concordance analysis.* We collected TOEFL iBT and IELTS Academic score reports but will include analyses of only the IELTS Academic results in this report (although see the [DET scores page](#) for the most up-to-date concordances between the DET and alternative ELP assessments). We continued to collect TOEFL iBT score reports following this study and will publish a report on the TOEFL concordance study at a later date. Test takers were included in the final concordance only if the following held: 1) the DET must be a certified test session taken after April 3, 2024[†] and before September 30, 2024; 2) the IELTS test date must be within 90 days of the DET test date; and 3) DET attempt must be one of the first seven test attempts for that test taker.

After collecting data, we constructed concordance tables for each of the SWRL subscores. To construct a concordance table, we chose pairs of scores to include, removed bivariate outliers, performed log-linear pre-smoothing and population weighting, and built the final concordance table using standard equating methods. Note that the distribution of DET scores in the concordance data does not match the marginal distribution of DET scores (as shown in [Figure 1](#)). The weighting step was implemented to adjust the sample weights so that the weighted distribution of DET scores in the concordance data would match the population distribution. Moreover, we compared kernel with equipercentile equating (see [Figure 3](#) in the Appendix) but only report results from the kernel equating procedure due to consistently lower and more stable bootstrapped standard errors across all conditions and subscores for kernel equating. To check the robustness of the final concordance, we systematically varied each of the above steps, which resulted in 12 total conditions to include in a sensitivity analysis. Additional specifics about the conditions that we varied in the study are as follows:

1. **Disambiguation:** If test takers have taken multiple DET and/or IELTS tests and more than one pair of these tests satisfies the inclusion rules listed above, choose the final DET and/or IELTS test to include that has the largest score value (“highest”) or the score corresponding to the latest attempt in the dataset used to conduct the concordance (“recent”).

*[Cardwell et al. \(2023\)](#) included both self-reported and official score reports in the concordance analysis and evaluated sensitivity of the final concordance map to how those data were combined.

[†]The DET began reporting subscores to test takers on July 1, 2024. However, these scores were calculated under-the-hood and available for analysis from April 3, 2024.

2. **Outlier:** Remove pairs from the data if the difference between the DET and IELTS tests are greater than 2 or 3 standard deviations (“2” and “3”) or keep all test takers in the data regardless of how dissimilar the DET and IELTS scores (“none”). The standard deviation difference is calculated as $|z_{\text{DET}} - z_{\text{IELTS}}|$, with $z_{\text{DET}} = \frac{x_{\text{DET}} - \bar{x}_{\text{DET}}}{s_{\text{DET}}}$, $z_{\text{IELTS}} = \frac{x_{\text{IELTS}} - \bar{x}_{\text{IELTS}}}{s_{\text{IELTS}}}$, where \bar{x}_T and s_T are the mean and standard deviation, respectively, of test T in the concordance data.
3. **Smooth Before:** We applied log-linear pre-smoothing (Davies et al., 2004) either on the raw data (“weighting”) or between weighting the data to match the population of DET test takers and the final equating step (“equating”).

Table 1 summarizes the conditions of the sensitivity analysis.

Table 1. Conditions of the concordance study sensitivity analysis.

Condition	Disambiguation	Outlier	Smooth Before
1	highest	2	equating
2	recent	2	equating
3	highest	3	equating
4	recent	3	equating
5	highest	None	equating
6	recent	None	equating
7	highest	2	weighting
8	recent	2	weighting
9	highest	3	weighting
10	recent	3	weighting
11	highest	None	weighting
12	recent	None	weighting

3 Results

The final concordance included a total of 1,942 test takers with DET–IELTS paired scores. Of these test takers, 591 had taken the IELTS after the corresponding DET test, and 1,351 had taken the IELTS before the corresponding DET test.

3.1 Descriptive Statistics

As described in the previous section, the final concordance dataset consists of 1,942 pairs of test sessions. This number contrasts with hundreds of thousands of test takers who have received a certified score on the DET in the same time period. We expect that the concordance dataset is sufficiently representative of the full DET population for the concordance to generalize to all DET test takers; most differences between the two datasets are accounted for and adjusted by the weighting procedure. Table 2 displays the descriptive statistics for the overall DET test-taker population as contrasted with the sample used to estimate the concordance. The DET scores are slightly higher with a narrower spread in the concordance data. These differences are notable for the speaking and writing scores, which have DET scores in the concordance data almost 3 points higher than the general population. The IELTS scores have similar means across the score types but with a smaller standard deviation for the speaking and writing scores. This result is nearly identical to those found in Ikeda et al. (2025). The correlation between the DET and IELTS is similar to but slightly lower than the subscore correlations reported in the most recent concordance between TOEFL and IELTS (Ikeda et al., 2025). These subscore correlations are typically between .5-.7, which is consistent with prior research on correlations between ELP tests.

Table 2. Descriptive statistics of the DET and IELTS in the full DET sample (DET (All) column) as well as the paired sample used to construct the final concordance tables (all other numeric). For the DET and IELTS-specific columns, values are reported as ‘mean (standard deviation)’ of the respective scores.

Score Type	DET (All)	DET (Paired)	IELTS Academic (Paired)	DET–IELTS Correlation
Overall	110.74 (22.02)	113.46 (16.16)	6.13 (0.77)	0.73
Speaking	109.84 (24.24)	113.63 (18.16)	6.06 (0.84)	0.68
Writing	109.78 (24.17)	113.45 (17.97)	5.86 (0.60)	0.54
Reading	110.21 (23.52)	111.97 (18.87)	6.08 (1.12)	0.53
Listening	110.60 (24.32)	112.25 (19.26)	6.28 (1.12)	0.57

Figure 1 displays a comparison of the estimated densities of the DET score in final concordance dataset as well as the DET distribution of all test takers over the same timespan. As noted earlier, we weighted the paired data to reflect the marginal DET distribution for each score type prior to

estimating the concordance. The densities in Figure 1 suggest that the sample with reported IELTS scores is slightly higher proficiency, with a more peaked DET distribution, than the general DET population.

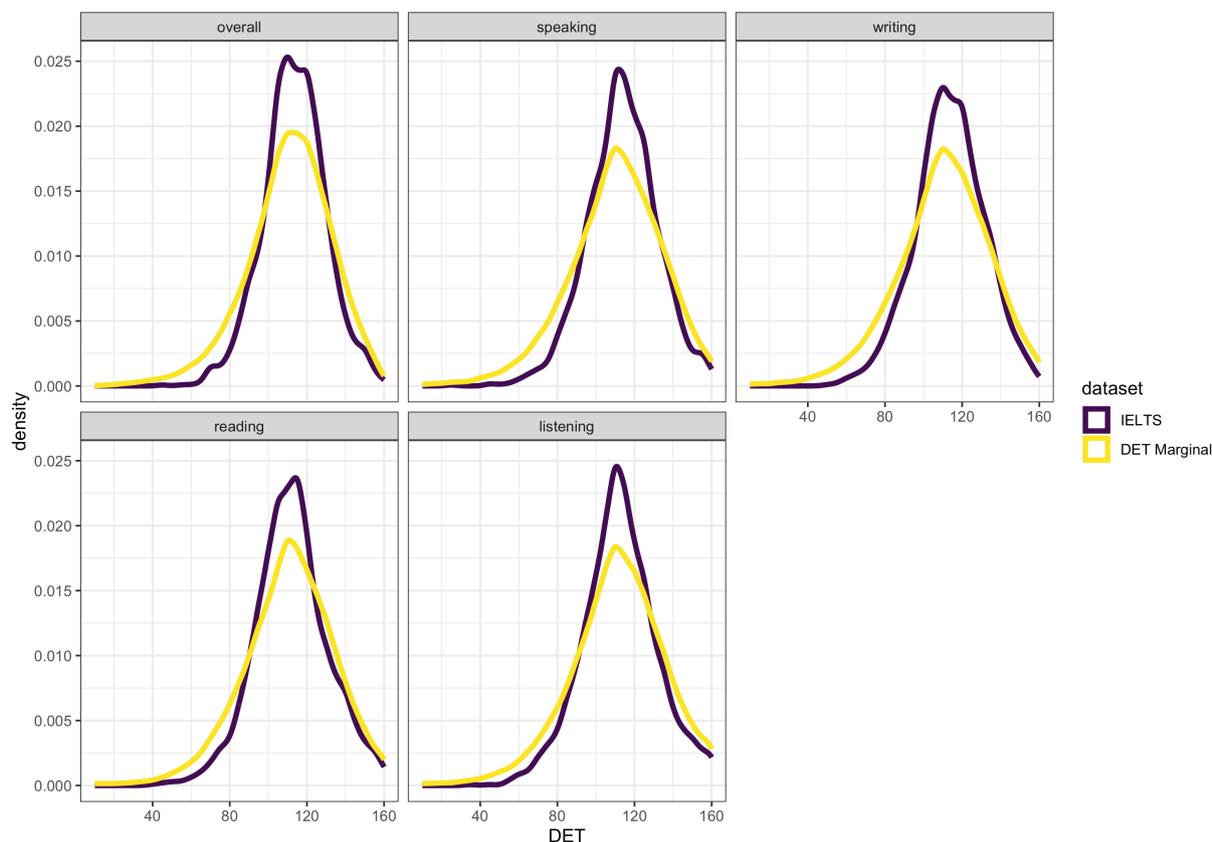


Figure 1. Comparison of estimated densities of DET scores within the final concordance dataset as well as the DET marginal dataset over the same time period. Different panels reflect different score types.

3.2 Concordance Comparisons

Figure 2 shows the concordance results across all conditions listed in Table 1 using kernel equating. As suggested earlier, the reading and listening concordances map the full range of DET scores to the full range of IELTS scores, whereas the speaking concordance only reaches an IELTS of 8.5, and the writing concordance reaches an IELTS of 7.5 or 8.0 (depending on whether weighting before smoothing or smoothing before weighting). The latter result was expected, as we received very few IELTS writing scores of 8.0 and no IELTS writing scores of either 8.5 or 9.0. A vertical line in each panel indicates the condition that matches the methodology of the overall concordance reported in Cardwell et al. (2023).

Similar to Cardwell et al. (2023), we chose condition 4 as the final concordance condition. Using the “highest” score resulted in the most up-to-date versions of both exams; setting an outlier threshold of 3 standard deviations allowed us to filter some high impact outliers but retain almost all of the data; and choosing smoothing between the weighting and equating steps resulted in the smoothing procedure only impacting the final step. Note that very few conditions affected the final concordance mapping, with the change in when weighting occurred having the most significant impact. These findings are consistent with those reported in Cardwell et al. (2023) for the overall score concordance.

3.3 Speaking

Table 3 displays the final concordance between the DET and IELTS Academic for the speaking subscore. Note that all DET scores between 10 and 60 have been collapsed into the same bucket and mapped to a similar range of IELTS scores. The highlighted IELTS range of upper A2 (4) to low C2 (8) maps to a DET range of between 65 and 155.

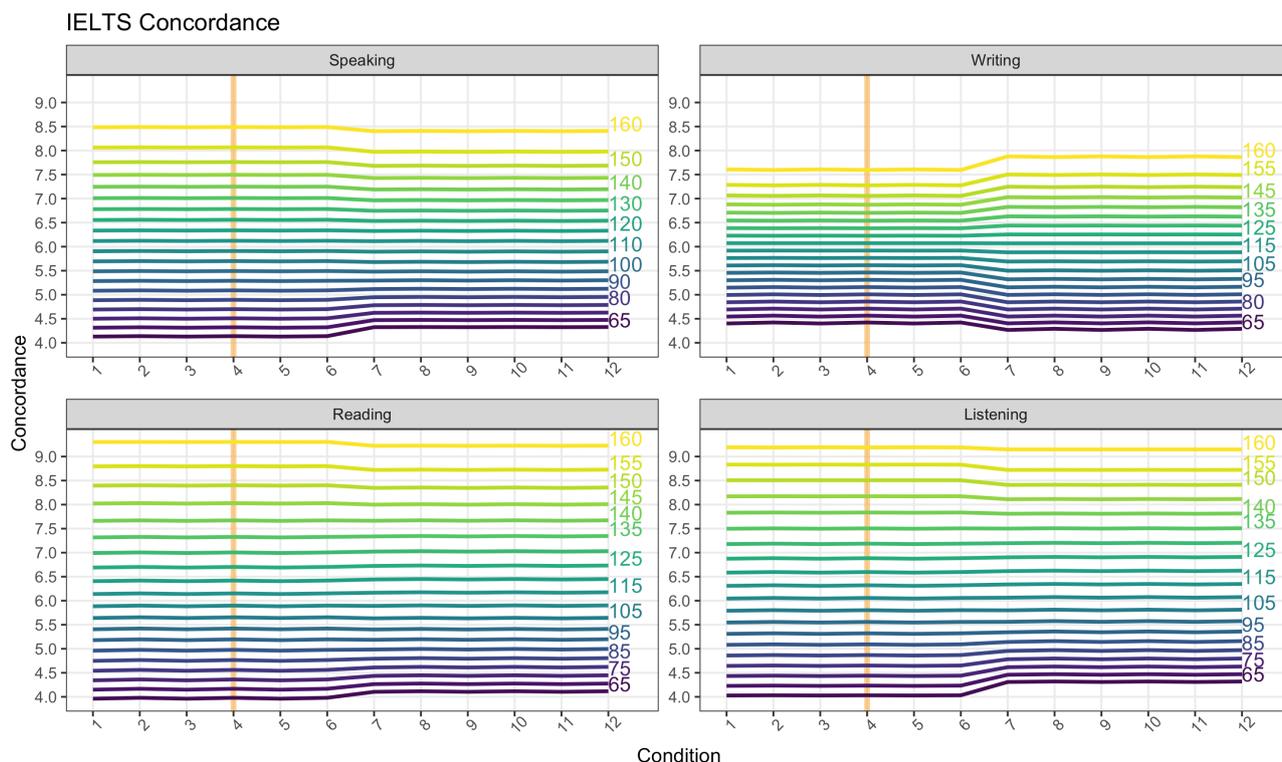


Figure 2. Concordance results across all conditions, with the *x*-axis indicating the condition number (see Table 1), the *y*-axis depicting the estimated unrounded IELTS score corresponding to a given DET score, and each line corresponding to a particular DET score. Different panels reflect different score types.

3.4 Writing

Table 4 displays the final concordance between the DET and IELTS Academic for the writing subscore. Note that all DET scores between 10 and 60 have been collapsed into the same bucket and mapped to a similar range of IELTS scores. The highlighted IELTS range of upper A2 (4) to low C2 (8) maps to a DET range of between 65 and 160.

As described above, we received only four IELTS writing scores of 8.0 and no IELTS writing scores of 8.5 or 9.0. This lack of IELTS writing data at the highest score points restricted the maximum value of the concordance and resulted in IELTS writing scores with a lower standard deviation than the other subscores. These results are consistent with those reported by Ikeda et al. (2025) when comparing the writing subscores of TOEFL iBT and IELTS Academic.

3.5 Reading

Table 5 displays the final concordance between the DET and IELTS Academic for the reading subscore. Note that all DET scores between 10 and 60 have been collapsed into the same bucket and mapped to a similar range of IELTS scores. The highlighted IELTS range of upper A2 (4) to low C2 (8) maps to a DET range of between 65 and 145.

3.6 Listening

Table 6 displays the final concordance between the DET and IELTS Academic for the listening subscore. Note that all DET scores between 10 and 60 have been collapsed into the same bucket and mapped to a similar range of IELTS scores. The highlighted IELTS range of upper A2 (4) to low C2 (8) maps to a DET range of between 65 and 145.

Table 3. Final concordance table for the speaking subscore using the selected method.

DET	IELTS Academic
160	8.5-9
150-155	8
145	7.5
130-140	7
120-125	6.5
110-115	6
95-105	5.5
85-90	5
70-80	4.5
65	4
10-60	1-4

Table 4. Final concordance table for the writing subscore using the selected method.

DET	IELTS Academic
160	8-9
155-160	7.5
145-150	7
130-140	6.5
110-125	6
95-105	5.5
80-90	5
65-75	4.5
10-60	1-4.5

Table 5. Final concordance table for the reading subscore using the selected method.

DET	IELTS Academic
155-160	9
150	8.5
145	8
135-140	7.5
130	7
120-125	6.5
110-115	6
100-105	5.5
85-95	5
75-80	4.5
65-70	4
10-60	1-4

Table 6. Final concordance table for the listening subscore using the selected method.

DET	IELTS Academic
155-160	9
150	8.5
140-145	8
135	7.5
125-130	7
115-120	6.5
105-110	6
95-100	5.5
85-90	5
75-80	4.5
65-70	4
10-60	1-4

3.7 Overall Score

In addition to the individual SWRL scores, we compared the concordance of the overall score in the current data with the overall score reported in Cardwell et al. (2023). Ideally the concordance between the overall scores would have not changed over the intervening two years. Table 7 shows the new overall score concordance using the chosen method between the DET and IELTS Academic overall scores.

Table 7. Concordance table for the overall score using the selected method.

DET	IELTS Academic
160	8.5
155	8.0
150	8.0
145	7.5
140	7.5
135	7.0
130	7.0
125	6.5
120	6.5
115	6.0
110	6.0
105	5.5
100	5.5
95	5.5
90	5.0
85	5.0
80	4.5
75	4.5
70	4.5
65	4.0

When comparing results from this concordance study to the prior one, only three score points had a change in the concordance map with the most recent study, all of those with a change of 0.5 IELTS score points, and none of a change larger than 0.5 IELTS score points. These changes are minimal and could be due to sampling variability and selection effects. Based on these results, we chose not to update the concordance between the DET and IELTS Academic overall scores from those reported in Cardwell et al. (2023).

We can also compare the actual overall concordance to the expected concordance based on aggregating individual SWRL concordances. That is, given a DET overall score and SWRL individual subscores, one can obtain a concordance value of an IELTS overall score in two different ways. The first strategy uses the DET overall score and the overall DET/IELTS concordance. The second uses each DET SWRL score and its associated DET/IELTS concordance to obtain concordance IELTS SWRL scores and then averages those scores. Ideally, these two methods would yield nearly identical IELTS overall scores, and that is what we find. Table 8 displays the actual and implied mapping of DET overall score to IELTS Academic overall score when aggregating the SWRL mappings. Based on Table 8, the differences between the actual and implied mappings are small (with a maximum absolute difference of 0.375 and one difference greater than 0.25 in absolute value), providing evidence that test takers and institutions would reach similar conclusions when comparing the new overall concordance to the 2022 overall concordance as well as the direct concordance mapping of overall score with the implied concordance mapping if taking the average of the individual SWRL mappings.

4 Conclusion

Concordance studies can be included as part of an external validation procedure. Tests that relate to each other should have stable mappings across successive concordances. Moreover, part of the process of constructing concordances (at least those that follow the recommendations from Knoch & Fan, 2024) is to assess the similarity of the tests and the stability of the inter-test relationships. This concordance study was designed, in part, to validate the overall concordance table described by Cardwell et al. (2023) and posted on the DET website. And when comparing the analysis reported in Cardwell et al. (2023) with that reported here, one only finds three differences each of only 0.5 IELTS points. These concordances are stable even though the DET has included new task types (such as Interactive Listening, Vocabulary in Context, and Interactive Writing) (Cardwell et al., 2024) and a new scoring model (Nydyck & Lockwood, 2024) in the interim.

In addition to validating the overall DET–IELTS concordance table, we provided concordance tables comparing DET and IELTS SWRL subscores. These subscore concordance tables will help institutions arrive at similar decisions regardless of the test that a test taker takes. To adequately

Table 8. Concordance table for the overall score using the selected method and compared to the mapping if aggregating concordances across S-W-R-L components.

DET	IELTS Academic (Actual)	IELTS Academic (SWRL Average)
160	8.5	8.62
155	8.0	8.38
150	8.0	8.00
145	7.5	7.62
140	7.5	7.25
135	7.0	7.12
130	7.0	6.88
125	6.5	6.50
120	6.5	6.38
115	6.0	6.12
110	6.0	6.00
105	5.5	5.62
100	5.5	5.50
95	5.5	5.38
90	5.0	5.00
85	5.0	5.00
80	4.5	4.62
75	4.5	4.50
70	4.5	4.25
65	4.0	4.12

conduct a concordance study, researchers must obtain sufficient test scores (across a variety of demographic groups and testing conditions) that reflect the typical range of test-taker behavior. All aforementioned concordance studies (Cardwell et al., 2023; Ikeda et al., 2025; Yu, 2021) struggled with obtaining sufficient sample sizes. Moreover, organizations that provide high-stakes assessments have little incentive to cooperate when creating concordances unless those organizations are non-competing. To conduct adequate sensitivity analyses of the final concordance, one must have a sufficiently large and representative sample across a variety of demographic groups who took pairs of tests in various orders and with different lengths of time between them. Finally, as concordances are between tests that are non-parallel, they are often constructed reluctantly and with various cautions and caveats (see Pommerich, 2007). Concordances allow users to make similar but not identical decisions, as the correlations between tests are often not high enough to treat the decisions as interchangeable.

As the DET gains popularity, fewer people take both the DET and alternative ELP tests, making it much more difficult to find the few test takers who continue to take multiple assessments. This group of test takers who both take multiple assessments and report official IELTS scores to DET researchers for a concordance study are self-selected and possibly different from the general group of DET test takers. As shown in Figure 1, this self-selected group has fewer scores at the lower and upper extremes for the DET. Moreover, as presented in Table 2, this group has a much narrower range of IELTS speaking and writing scores than reading and listening scores.* The lack of test takers at extreme score points means that the concordance at those points is less stable than at other score points (see Figure 3 in the Appendix). We would caution about making equivalence claims for decisions at these score points without additional data.

As described by Pommerich (2007), concordances are beneficial to users, as they “give students and schools more flexibility in their testing options ... reduce testing costs and allow students more choices” (p. 212). Although decisions across different tests can never be identical, we hope that the concordances can help institutions make fair and equitable decisions.

*IELTS does not post full score distributions on their website. Therefore, we cannot determine whether the score distributions that we obtained for IELTS are similar to those from a typical sample of IELTS test takers.

5 References

- Cardwell, R. L., Naismith, B., LaFlair, G. T., & Nydick, S. W. (2024). *Duolingo english test: Technical manual*. https://duolingo-papers.s3.amazonaws.com/other/technical_manual.pdf
- Cardwell, R. L., Nydick, S. W., Lockwood, J. R., & Von Davier, A. A. (2023). Practical considerations when building concordances between English tests. *Language Testing*, 02655322231195027. <https://doi.org/10.1177/02655322231195027>
- Davier, A. A. von, Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating*. Springer New York. <https://doi.org/10.1007/b97446>
- ETS. (2012). *The official guide to the TOEFL® test*. Fourth edition. New York, NY : McGraw-Hill, 2012.
- 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* (p. 24). <https://ielts.org/researchers/our-research/research-reports>
- Knoch, U., & Fan, J. (2024). Test score comparison tables: How well are they serving test users? *Language Testing*, 02655322241239348. <https://doi.org/10.1177/02655322241239348>
- LaFlair, G. T. (2020). *Duolingo English Test: subscores*. <https://duolingo-papers.s3.amazonaws.com/reports/subscore-whitepaper.pdf>
- Nydick, S. W., & Lockwood, J. R. (2024). *An overview of Duolingo English Test administration and scoring*. <https://doi.org/10.46999/RCXB1889>
- O’Sullivan, B. (2018). *IELTS (international english language testing system)* (pp. 1–8). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118784235.eelt0359>
- Pommerich, M. (2007). *Concordance: the good, the bad, and the ugly* (N. J. Dorans, M. Pommerich, & P. W. Holland, Eds.; pp. 200–216). Springer New York. http://link.springer.com/10.1007/978-0-387-49771-6_11
- Yu, G. (2021). *IELTS academic and PTE-academic: Degrees of similarity*. (pp. 7–41).

6 Appendices

6.1 Bootstrapping

As described in the paper, our concordance process involved several manipulated factors. To determine the standard error of equating (SEE), we implemented a bootstrap procedure identical to that implemented in Cardwell et al. (2023). See an appendix in that paper for more information about the logic of the bootstrap procedure. As shown in Figure 3, kernel equating resulted in less jagged SEE functions than equipercentile equating. This result was expected due to the limited number of score points for IELTS and provides additional evidence for using kernel equating rather than equipercentile equating in this case.

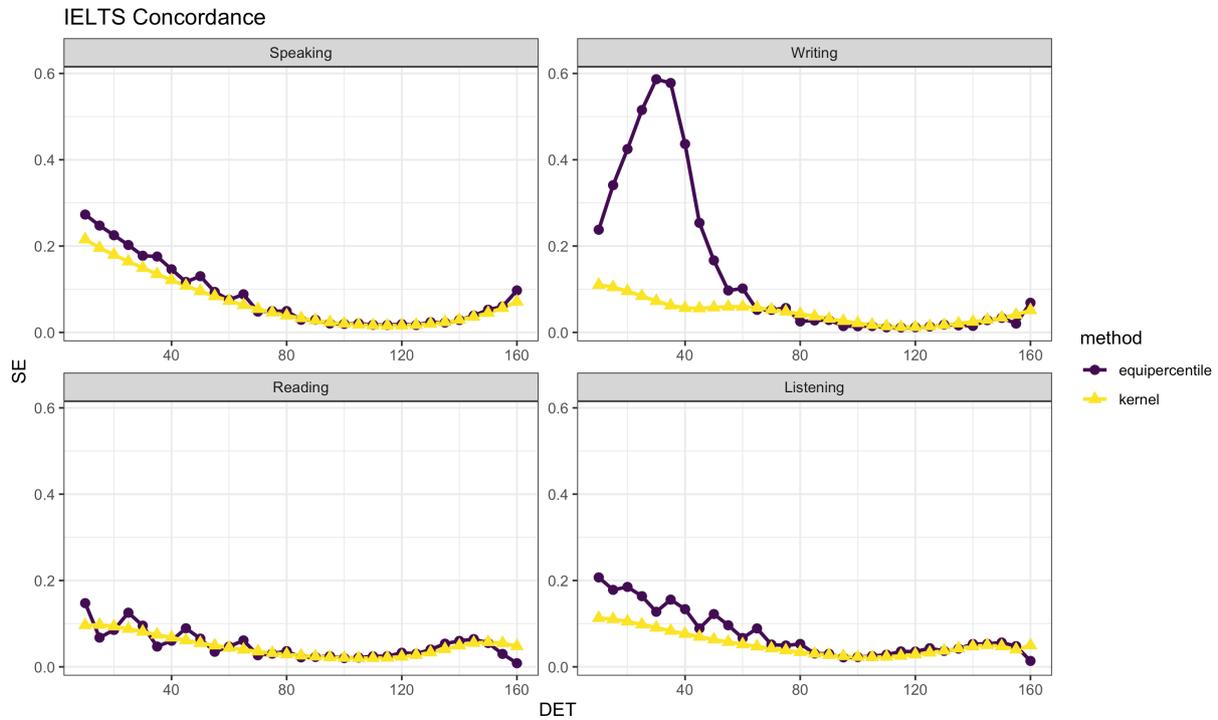


Figure 3. Bootstrap standard errors of concordance condition 4 from Table 1.