Pearson
Test of English

Sarah R. Hughes | Dr. Rose Clesham

Establishing score concordance between the enhanced PTE Academic and IELTS Academic

Concordance Analysis

Contents

Pearson

July 2025

Executive Summary

This paper reports on the results of a concordance study conducted in 2024 between the enhanced PTE Academic and IELTS Academic. It improves on previous studies, offering more detailed and rigorous results that are vital for academic, professional, and migration decisions.

The enhanced PTE Academic in large part retains its essential structure. It tests all four language skills using a combination of single and integrated skills. No question types have been removed; however, two new extended speaking tasks have been added, alongside enhanced scoring procedures.

The study involved 1,522 participants who took both the enhanced PTE Academic and IELTS Academic tests within 90 days and in a reasonably counterbalanced testing order. Participants were recruited from India, China, Australia, and the United Kingdom, representing 54 nationalities, 43 languages, and a wide range of language skills.

The tests were taken under strict conditions to ensure accurate results. Participants were given preparation materials for both tests to ensure they were familiar with them. The analysis showed that the sample of participants was representative of the global population of PTE Academic test takers.

The concordance study followed in detail established good practice principles for concordance studies, ensuring robust score comparisons on all four skill scores as well as the overall test scores. The concordance tables were produced through rigorous analysis using established methodologies. Full methodological details and statistical information are provided in the report.

Overall, this concordance study is more thorough than previous studies, with a larger sample size and more detailed reporting. The results provide updated and reliable score comparisons for the enhanced PTE Academic and IELTS Academic, helping institutions and governments make informed decisions about language proficiency. Pearson understands the importance of regular score concordance studies to support accurate and relevant comparisons between English language proficiency tests.

Introduction

Purpose of this study

This paper reports on the results of a concordance study conducted in 2024 between the enhanced PTE Academic test and the IELTS Academic test.

Concordance studies between high-stakes English language tests are essential in order to support the interpretation and standards of test scores from different tests that are used for common purposes. PTE Academic and IELTS Academic are both used for academic, professional, and economic migration around the world. This concordance study supports the interpretation of the relationship between the scores of these two tests in this context. These high-stakes tests scores carry significant currency, and it is therefore essential that test takers and test score users understand the direct relationship between test scores.

An initial concordance between PTE Academic and IELTS Academic was established by Zheng & De Jong (2011). This was followed up by an updated concordance study in 2020 (Clesham & Hughes, 2020).

Updating concordance score comparisons over time is necessary as several test variables can periodically change. These include testing populations, test preparation practices, test familiarity and test purposes. While these reasons alone may have necessitated an updated concordance study between PTE Academic and IELTS Academic, the primary reason for conducting a new concordance study in 2024 was the introduction of enhancements to PTE Academic in July 2025.

The enhanced PTE Academic test in large part retains its essential structure. It tests all four language skills using a combination of single and integrated skills. No question types have been removed; however, two new extended speaking question types have been added, alongside enhanced scoring procedures.

The enhanced PTE Academic test was thoroughly field tested in early 2024. The enhanced test design and analyses included measures to maintain the alignment of the enhanced PTE Academic test to the Common European Framework of reference (Council of Europe. 2001), using on both standard setting and statistical linking procedures.

Therefore, the primary purpose of this 2024 concordance research study is to provide updated score concordance tables between the enhanced PTE Academic test and the IELTS Academic test. All the methodological steps of this concordance study follow the good practice principles developed by Knoch & Fan (2024) ensuring robust score comparisons on all four skill scores as well as the overall test scores. Table 1 shows the good practice principles along with page references to locate the evidence in the report related to each principle.

Table 1: *Good practice principles (Knoch & Fan, 2024) and location of evidence related to each principle*

Principle	Page Reference
Verify the two tests measure closely related constructs Ensure that test administration conditions are similar Maintain similar levels of reliability for the overall test and subsections	6-7
Establish strong correlations between the two tests	18
Ensure that the participant sample mirrors the test population of interest	12-15
Collect data on the participants' reasons for taking the tests	13-14
Ensure that the participants have comparable levels of test preparation and familiarity with the tests	10-11
Ensure that the data are based on official test score reports	9
Ensure that the sample size is adequate	9
Implement a counter-balanced design	9, 29-31
Ensure that the interval between test takers attempting both tests is sufficiently short	9
Conduct a population invariance study	32-37
Include descriptive statistics and correlation coefficients in the report	15-18
Describe statistical methods and procedures in sufficient detail	19
Report concordance results for both the overall test and subsections	20-23
Report the sample size and standard error at each score level	21-23
Alert test users to exercise caution in interpreting and using concordance results	23-24
Provide test user-focused guidelines and recommendations	20-24

Rationale for aligning PTE Academic and IELTS Academic

PTE Academic and IELTS Academic are both designed to be tests of language proficiency used in academic and professional high stakes entry and selection decisions. Both tests align to a common framework of language proficiency, the CEFR, using methodologies recommended by the manual *Relating Language Examinations to the Common European Framework of Reference for Languages* (Council of Europe, 2009). Because each test claims an alignment to the CEFR, a level of commonality is intended in the score interpretation. However, it should be noted that alignments to frameworks are not empirical and rely on expert judgement exercises. As such, test-to-framework alignments can vary based on the chosen methodologies and can be influenced by unconscious bias and heuristics of the human experts involved. Concordance studies are important additional evidence that offer a clearer, empirical view of the relationship between scores on distinct tests. This concordance study was a data-led empirical study on test-to-test score comparisons.

PTE Academic and IELTS Academic are purposed to measure the same populations in terms of the four language skills in the context of selection for study or professional employment. Therefore, conducting empirical concordance studies on corresponding test scores is essentially pragmatic. Both tests' scores are actively used in the same high-stakes decisions, and so an understanding of their relative meaning is a necessity for test takers and decision-makers alike. While it is a necessary comparison, it should also be considered to what extent it is a meaningful comparison. To what extent is it appropriate to align the scores of these tests? Scores from different tests should only be linked if the tests share demonstrably similar characteristics and statistical properties (Kolen & Brennan, 2014).

Test design and construct comparability

Prior research has established the appropriacy of linking PTE Academic and IELTS Academic (Zheng & De Jong, 2011), however, given the changes to the enhanced PTE Academic, a new external study was commissioned to compare the construct of the enhanced PTE Academic and IELTS Academic. The construct comparison was undertaken by independent researchers at the University of Bedfordshire Centre for Research in English Language Learning and Assessment (CRELLA). An excerpt from the executive summary of the research report is shown below.

[The construct comparison research carried out by CRELLA] presents a comprehensive analysis of the enhanced Pearson Test of English Academic (PTE Academic), examining its construct and 22 question types through the socio-cognitive framework (Weir, 2005). Test specifications, sample materials, and test-taker performances provided by Pearson were scrutinised and matched with publicly available information and previous research about IELTS Academic.

PTE Academic was found to be a vigorous test which meets or exceeds the standards expected of high-stakes academic language tests in terms of task design (context validity), cognitive processes required of test takers (cognitive validity) and the levels of test taker performance (scoring dimension). Further, its use of authentic input materials sets the test firmly in the academic domain. Its emphasis on skill integration is congruent with contemporary research which illustrates the interconnected nature of language skills.

Comparing PTE Academic and IELTS Academic, both tests are intended for higher education contexts and cover CEFR A (Basic User), B (Independent User), and C (Proficient User). For Reading, PTE Academic emphasizes diversity in passage range, whereas IELTS Academic offers greater variety in question types. Both engage major cognitive processes involved in reading, including word recognition, lexical access, syntactic parsing, and inferencing, but neither assesses intertextual reading across multiple texts. Both cover a range of reading types, providing comprehensive evaluations of reading skills. PTE Academic has a temporally shorter reading section than IELTS due to its sampling of more but shorter question types, including the often under-represented academic expeditious reading. The reliability and validity of the shorter section should be confirmed with further field and concordance testing.

Though both tests assess similar listening processes, PTE Academic focuses exclusively on academic and professional listening contexts, while IELTS Listening—overlapping with IELTS General Training—includes the personal domain relevant to more general conversations. This distinction reflects differences in priorities between academic-specific assessment and broader language use domains. Comparison of Speaking and Writing between IELTS Academic and PTE Academic is not straightforward due to difference in the independent and integrated approaches of their test design. PTE Academic's integrated design combines Speaking and Writing into a single section—IELTS treats these skills as distinct. PTE Academic Speaking tasks require test-takers to synthesize spoken and written inputs (e.g., Describe Image; Repeat Sentence) which evaluate skills such as pronunciation, fluency, and the ability to process and convey information. Integrated tasks (e.g., Retell Lecture; Answer Short Question) mimic real-world academic tasks and interactions, motivating the longer test time allocated to Speaking than in IELTS. For Writing, by incorporating both short and extended responses, PTE Academic engages a wide range of abilities, including concise academic summaries to detailed essay tasks. Writing tasks emphasize formal register, accurate grammar, and clarity, reflecting real-world academic requirements.

Given the substantial overlap between PTE Academic and IELTS Academic, including the academic English domains targeted, the cognitive processes elicited, and the level of test taker performance expected, the two tests were found to have overall comparable constructs.

Full detail of the construct comparison analysis can be found in Establishing score concordance between the enhanced PTE Academic and IELTS Academic – Background: Construct Comparison (2025).

Statistical properties

Reliability is a necessary but insufficient criterion for any assessment to be valid and accurate. In a concordance study, the combined measurement error of both tests will affect the precision and therefore usefulness of results. As such, an investigation into assessment reliabilities and measurement error is a necessary first step for any concordance study. Table 2 shows the reliability statistics for the updated PTE Academic and the publicly reported reliability statistics for IELTS Academic.

IELTS Academic reports reliability as ranges of Cronbach's alpha internal reliability estimates for receptive skills and inter-rater reliability statistics for productive skills. IELTS states that their composite reliability estimate was calculated following Feldt & Brennan (1989) (*IELTS Test Performance Data 2023-2024*, 2024).

PTE Academic reliability estimates for all scores are based on the person reliability of the psychometric models used in scoring. Note that the statistic Winsteps reports for person reliability is the mathematical equivalent of Cronbach's alpha in the context of psychometrically scored tests (Linacre, 2024). Though a variety of different reliability statistics are reported, they are all robust, exceeding 0.8.

Table 2: *Reliability estimates for PTE Academic and IELTS Academic*

	Overall	Listening	Reading	Speaking	Writing
PTE Academic	0.93	0.89	0.82	0.85	0.82
IELTS Academic	0.97	0.91-0.94	0.85-0.92	0.88	0.95

Table 3 shows the Standard Error of Measurement (SEM) for PTE Academic and the publicly available information on IELTS SEM. IELTS does not provide SEM statistics for productive skills in their public test performance report (2024). Ideally test providers would report SEM information for all scores, and at different points across the score scales, referred to as Conditional Error of Measurement (CEM). See Appendix B for the enhanced PTE Academic's CEM values.

Table 3: *Standard Error of Measurement (SEM) for PTE Academic and IELTS Academic*

	Overall	Listening	Reading	Speaking	Writing
PTE Academic	3.09	4.22	5.19	5.75	4.63
IELTS Academic	0.17	0.38	0.41	-	-

Note. PTE Academic is scored on a scale of 10-90 and IELTS Academic is scored on a scale of 1-9.

Methodology

Study design overview

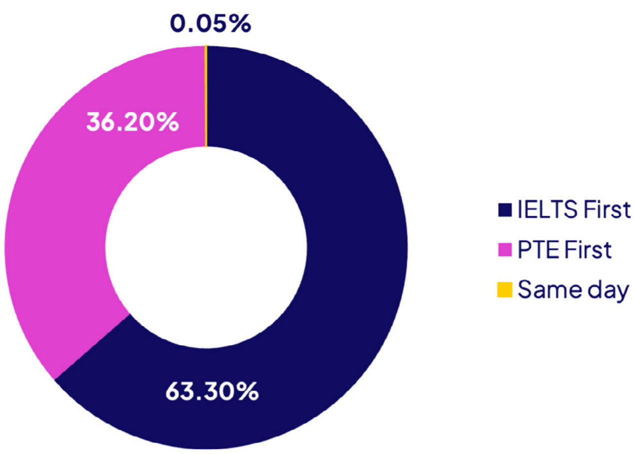
The concordance study was designed to adhere to good practice principles in data collection (Knoch & Fan, 2024). The concordance study used a single group design, with all participants sitting the enhanced PTE Academic and providing a .pdf file of their official IELTS Academic score report. All score reports were manually checked by the research team to verify the authenticity of test taker information.

We aimed to collect a sample exceeding 1,000 test takers who had taken both tests within 90 days of each other. Test takers who provided a score report for an IELTS test taken more than 90 days before or after their PTE Academic test were excluded from the final data analysis. On average, the time between the two tests was 18 days.

We aimed for the sample to be counterbalanced, so that equal numbers of people took an IELTS test first and a PTE test first. An equally counterbalanced sample exceeding the minimum sample size requirement was collected (n=1,109), with 551 PTE-first test takers, 551 IELTS-first test takers, and 7 same-day test takers. In addition to this, we were able to collect further concordance data from test takers who had already taken an IELTS Academic test and only needed to sit the enhanced PTE Academic test (n=413).

An analysis was conducted to determine if the additional data from IELTS-first test takers would be beneficial to the concordance analysis. This analysis compared concordance tables based on a completely counterbalanced sample with those based on a sample with additional data. The investigation found negligible difference between the concordance tables and no significant order effect in the scoring data. Therefore, additional data were considered beneficial to the concordance analysis, and the full sample of 1,522 test takers was retained. Figure 1 shows the test order balance of the full sample, including additional data. *See Appendix A – Test order effect analysis* for more information.

Figure 1: *Concordance study testing order*



Recruitment and data collection

The concordance study data collection took place in India, China, Australia, and the United Kingdom from September to November 2024. India and China were a focus of recruitment, as these two countries account for the majority of PTE's global test taking population. Efforts were made to recruit test takers from across a representative range of proficiency. In India and China, additional recruitment support was provided by language schools and research partners to ensure participation in the study would be accessible to test takers with lower language proficiency. Australia was also a focus of recruitment, targeting test takers who had previously taken PTE and successfully immigrated to Australia. Additionally, a small number of participants in the UK were included to ensure sufficient representation of participants whose primary language is English. Recruitment in Australia and the UK was direct-to-test-taker, as many test takers in these locations were capable of navigating participation in the study independently.

Test structure and preparation

To conduct a concordance study between the operational IELTS Academic test and the enhanced PTE Academic test under development, test takers sat a field test administration of the enhanced PTE Academic under secure exam conditions in official test centres. The field test administration of the test included the exact test structure, question types, scoring criteria, and scoring model used by the enhanced PTE Academic test.

It is important that test takers are adequately prepared and put forth full effort in taking both tests so that valid comparisons in score attainment can be made. All participants were provided with free preparation materials for both tests. For IELTS, test takers were provided with free official IELTS preparation resources to familiarize themselves with the test. For the enhanced PTE Academic, a special offline practice test was created to illustrate the updated test structure and new question types. A survey, which included questions about test familiarity and test preparation, was sent to all 1,522 participants in the study, with 663 participants responding. Figure 2 shows the levels of familiarity participants had with each test prior to their participation in the concordance study. Once recruited, participants were required to engage with familiarisation materials for both tests and to complete both tests with full effort. In total, 95% of participants reported preparing for IELTS and 96% reported preparing for the enhanced PTE Academic test. Figure 3 shows the preparation methods used by participants in the study.

Figure 2: *Test familiarity prior to participation in the concordance study (n=663)*

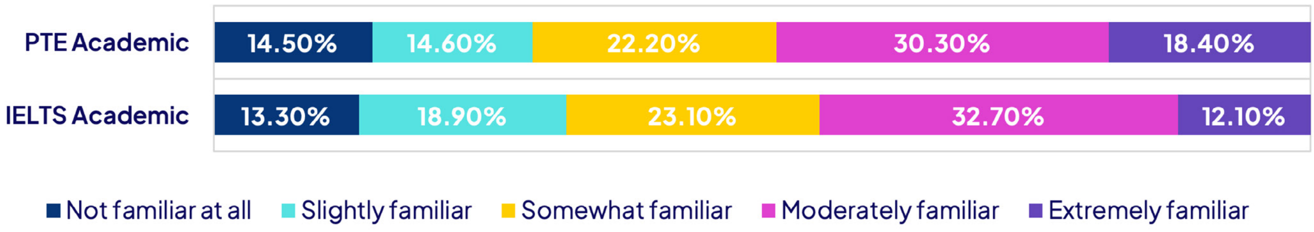
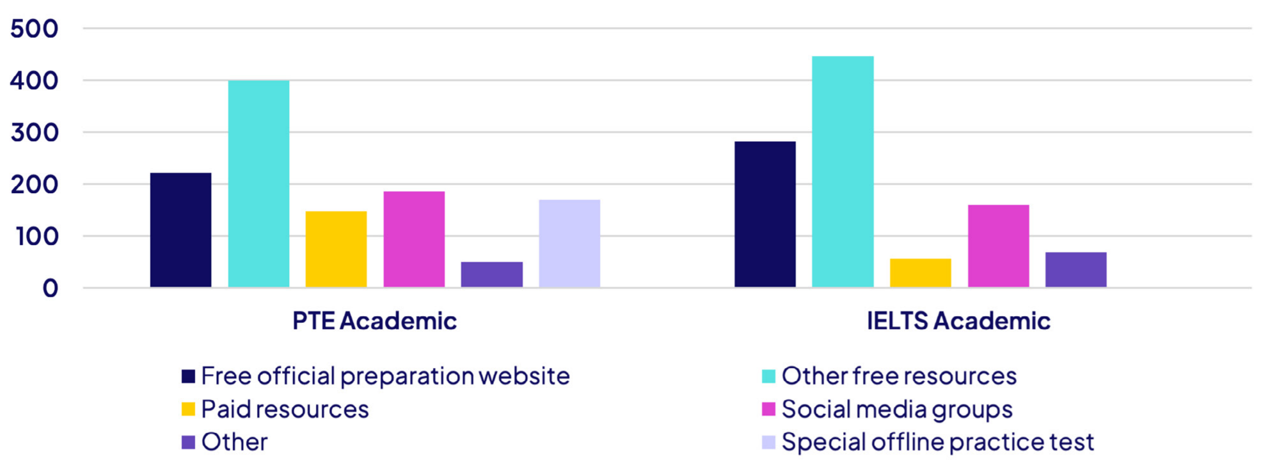


Figure 3: *Methods of test preparation used by study participants (n=663)*



Overall, test takers were largely familiar with both tests prior to participation in the study, as they were either currently engaged in preparation for a test or had taken a test as part of their immigration journey. Participants made use of free preparation resources, including the official preparation materials and the special offline practice test created for the concordance study. Some participants engaged in paid preparation, though this is to be expected, as many participants will have engaged in preparation for one or both of the tests prior to entry into the study (in particular, participants who had successfully immigrated to their destination country and participants recruited through language schools). The findings of the survey indicate that test takers were similarly familiar with both tests and engaged in preparation activities to familiarize themselves with both tests.

Analysis

To link two distinct tests of similar purpose and design, the representativeness of the sample and the strength of the empirical relationship between the test scores must also be established. The following analyses consider the representativeness of the sample and strength of the relationship between test scores.

Sample representativeness

Demographics

For a concordance study to be generalizable and useful in decision-making, the concordance sample must be representative of the population about which decisions are being made – in this case, the global test taker population for PTE Academic. Figures 4-6 and Tables 4-5 compare the concordance study representation of these characteristics to the live testing population of PTE's global test takers in 2024 in terms of gender, age, reason for testing, nationality, and language spoken at home.

Figure 4: *Comparing concordance study sample and PTE's 2024 global population gender*

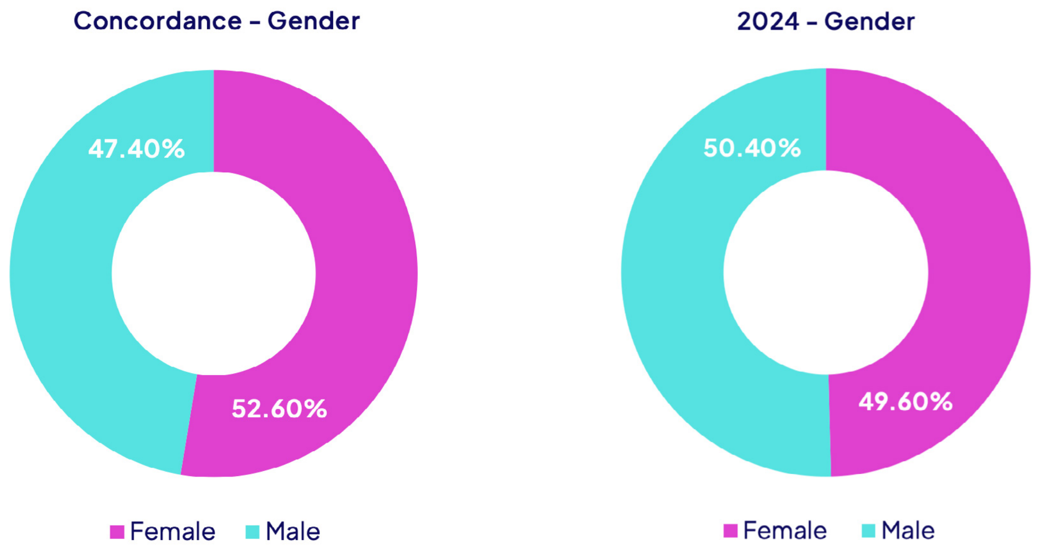


Figure 5: *Comparing concordance study sample and PTE's 2024 global population gender*

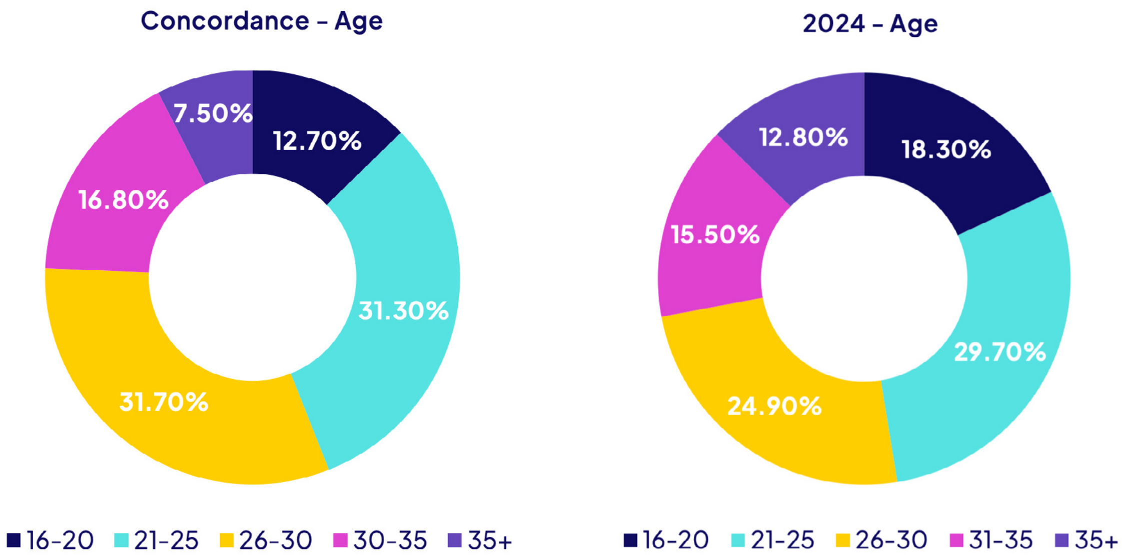
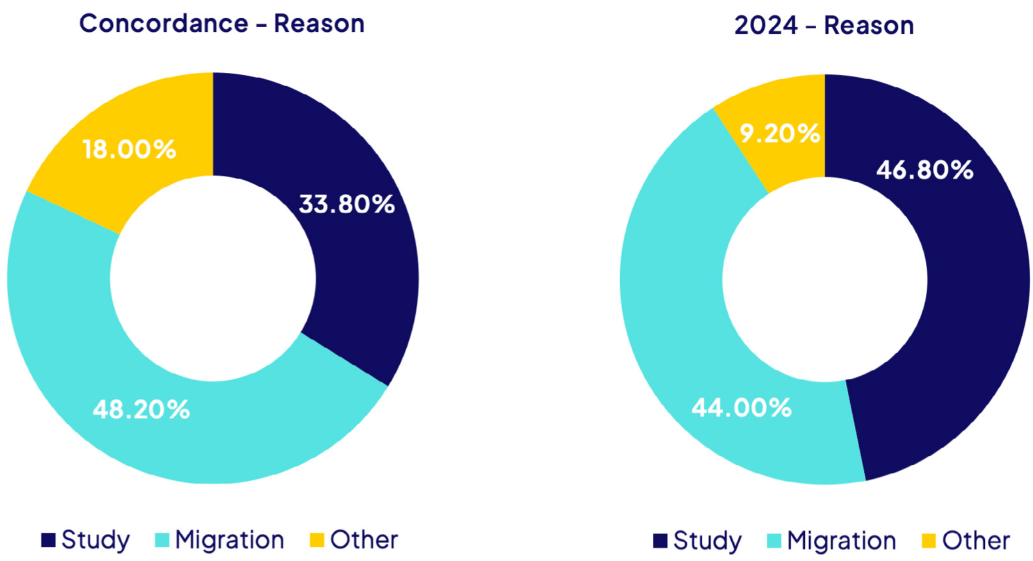


Figure 6: *Comparing concordance study sample and PTE's 2024 global population reason for testing*



The split of gender and age groups found in the concordance study is nearly identical to the live testing population. While registering to take a PTE Academic test, participants are asked to give their reason for taking a test. Approximately half of both the concordance sample and live testing population are testing for migration purposes, and a large portion for study purposes. The concordance sample includes a slightly larger group of test takers testing for other purposes. This can be attributed to the fact that more test takers made use of the open response option for "other" on the concordance survey, when their responses could have been classified into either migration or study. The concordance study sample's reasons for testing reflect the range of reasons for testing found in the live population.

In total, 54 nationalities and 43 languages are represented in the concordance study sample.

The tables below show the top five most common nationalities (Table 4) and languages (Tables 5) in PTE's live test taking population, alongside the percentage of the concordance sample these represent.

Table 4: *Top 5 PTE 2024 global test taker nationalities and their representation in the concordance study sample*

Rank	Countries	Percentage of 2024	Percentage of concordance
1	India	36.9%	21.2%
2	China	12.4%	28.4%
3	Nepal	11.4%	3.2%
4	Philippines	6.7%	1.3%
5	Pakistan	5.3%	2.6%
-	Other	27.3%	43.3%

Table 5: *Top 5 PTE 2024 global test taker languages and their representation in the concordance study sample*

Rank	Language	Percentage of 2024	Percentage of concordance
1	Chinese-Mandarin	12.8%	41.9%
2	Hindi	12.1%	12.5%
3	Punjabi	12.0%	4.9%
4	English	11.6%	3.7%
5	Nepalese	12.1%	3.0%
-	Other	39.4%	34.0%

In both the concordance sample and the live testing population, the majority of test takers hold either Indian, Chinese, or Nepalese citizenship and speak languages associated with these countries. The concordance study sample has greater representation of Chinese test takers and Chinese-Mandarin speakers than is currently found in the live testing population. This pattern was also present in earlier field-testing research for the enhanced PTE Academic, and additional recruitment streams were introduced in the concordance study to increase representation of other nationalities and languages. However, the completion rate of Chinese test takers was higher than that of other demographic groups. Though Chinese citizens and Chinese-Mandarin speakers are over-represented in the sample, they have been retained in the sample for the concordance analysis. Given the breadth of nationalities and languages included in the sample, as well as the ample representation of the largest demographic groups for PTE's live testing population, there is negligible risk of under-representation of other key demographics. Further, Chinese test takers and Chinese-Mandarin speakers are growing demographics for PTE, and they rightly form a significant portion of the concordance sample. As shown in the figures above, the concordance sample of 1,522 participants is both large and representative of key demographics for PTE Academic's global test taking population.

Score distributions

Concordance samples should represent the full range of proficiency that each test claims to measure. To evaluate the representativeness of the concordance sample range of proficiency, the concordance sample score data can be compared to live test score data. The following analysis considers the concordance score distributions, means, and standard deviations in relation to operational scoring information where available. For PTE, operational scoring information is provided for the 2024 global test taking population. For IELTS, operational scoring information is based on the IELTS Test Statistics webpage 2023-2024 global test taker performance data (IELTS, 2024). These operational data sets are not directly comparable, as there is some variation in the time periods covered and statistics provided. However, they provide a useful reference point for the data collected throughout the concordance study.

Mean scores are useful in understanding how the overall ability of the concordance sample compares to the wider testing population. Table 6 shows the mean PTE and IELTS concordance sample scores in relation to a live testing population. For PTE Academic, the 2024 global testing population mean scores are provided. For IELTS, mean score data is only publicly available for subgroups, so mean score by gender is provided for the 2023-2024 IELTS Academic global testing population.

Table 6: *Comparing mean scores from concordance sample with live testing populations for PTE and IELTS*

	IELTS Mean Scores		PTE Mean Scores	
	2023-2024	Concordance Sample	2024	Concordance Sample
Listening	6.38 M / 6.38 F	7.03	59.16	60.60
Reading	6.24 M / 6.28 F	6.71	59.74	58.00
Speaking	5.89 M / 5.95 F	6.39	61.33	64.68
Writing	5.97 M / 6.05 F	6.23	59.93	58.98
Overall	6.18 M / 6.23 F	6.66	59.37	60.77

The mean IELTS scores in the concordance sample were 0.18-0.65 higher than the mean scores reported in the 2023-2024 test performance statistics. This indicates that the concordance sample was slightly more able than the global IELTS testing population. For PTE, the concordance sample mean scores are largely in line with the global testing population.

Score distribution information is useful in understand how scores are spread across each test's score range, giving an indication of the test's ability to differentiate levels of proficiency. Figure 7 shows score distribution information in terms of boxplots of PTE Academic scores for the 2024 global testing population alongside concordance sample scores for PTE Academic and IELTS Academic.

Figure 7: *Comparing PTE Academic 2024 scores concordance sample scores for PTE Academic and IELTS*

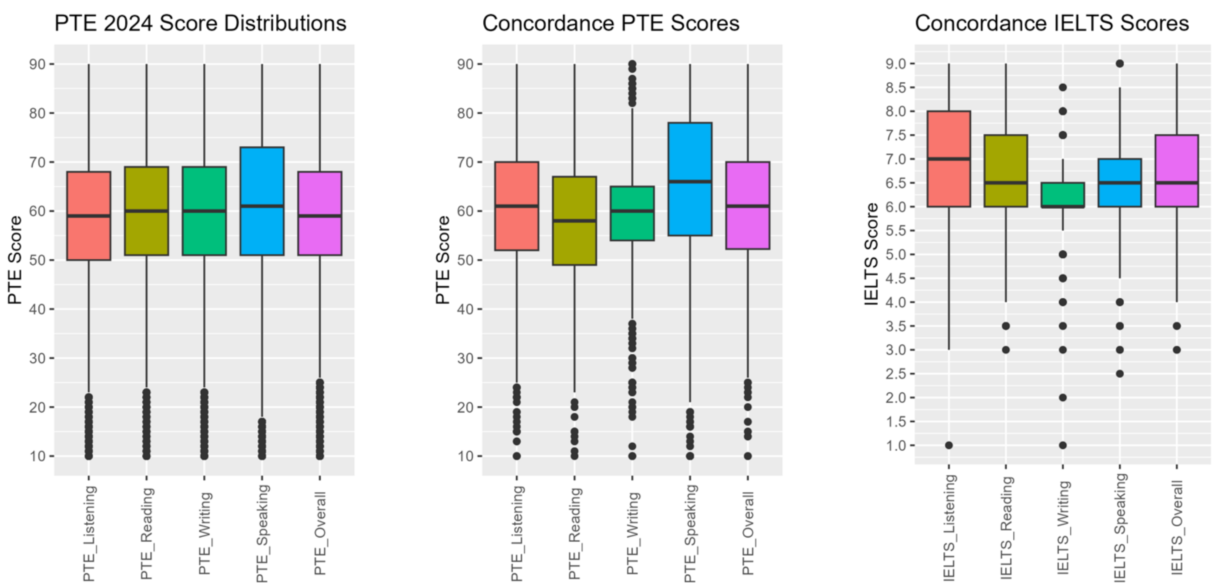


Figure 7 shows the PTE concordance score distributions have a similar median and interquartile range to the 2024 PTE global testing population. Figure 7 shows the IELTS score distribution for the concordance sample only, as the 2024 IELTS score distribution information is not publicly available.

Table 7 shows score distribution information as standard deviation in terms of score points and relative standard deviation in terms of percentage of each test's score scale. Note that there is no publicly available score information for standard deviation in the 2023-2024 IELTS global test taker performance data. However, standard deviations for Listening and Academic writing are available from 2022, and these are given in Table 7.

Table 7: *Comparing standard deviation and relative standard deviation from concordance with live testing populations*

	IELTS Standard Deviation SD (RSD)		PTE Standard Deviation SD (RSD)	
	2022	Concordance Sample	2024	Concordance Sample
Listening	1.35 (15%)	1.26 (14%)	14.33 (18%)	14.90 (18%)
Reading	1.26 (14%)	1.22 (14%)	14.13 (17%)	13.77 (17%)
Speaking	NA	0.90 (10%)	17.41 (21%)	18.01 (22%)
Writing	NA	0.65 (7%)	13.20 (16%)	11.32 (14%)
Overall	NA	0.89 (10%)	13.39 (17%)	13.67 (17%)

Taken together, Figure 7 and Table 7 show that the two tests make somewhat different use of their score scales. IELTS' relative standard deviations (7-14%) are consistently smaller than PTE Academic's (16-21%), indicating relatively less differentiation among concordance test takers. This effect is most noticeable for productive skills, where the relative standard deviation of concordance scores for IELTS is approximately half the relative standard deviation for PTE Academic. Because the IELTS standard deviations observed in the concordance study are in line with the publicly reported 2022 standard deviations for Listening and Reading, it is likely that the standard deviations for Speaking and Writing are representative as well. While this finding does not preclude concordance, it should be considered in the interpretation of concordance tables, particularly at the extremes of each scale.

Correlations

In addition to scoring data being reliable and representative, there must also be an empirical relationship between the two tests' scores to establish an alignment between them. Figure 8 shows the correlation matrix for PTE and IELTS scores in the concordance sample.

Figure 8: *IELTS vs PTE score correlation matrix*

	PTE Listening	PTE Reading	PTE Writing	PTE Speaking	PTE Overall
IELTS Listening	0.70	0.61	0.57	0.57	0.70
IELTS Reading		0.66	0.57	0.52	0.67
IELTS Writing			0.56	0.51	0.65
IELTS Speaking				0.60	0.71
IELTS Overall					0.78

There is a strong correlation (r=0.78) between overall scores on IELTS Academic and PTE Academic, with moderate to strong same-skill subscore correlations (r=0.56-0.70). This is an expected finding, particularly given PTE Academic's approach to integrated skills assessment, whereas IELTS assesses skills independently. Further, skill subscores are necessarily based on fewer score points and less test taker information than overall scores. This context is important in the interpretation of test scores and score concordance tables. Overall scores and score concordances will always provide the most reliable basis for decision-making. However, the strength of the skill subscore correlations shown in Figure 8 demonstrates that they are sufficiently strong to support linking.

Equipercentile linking

An equipercentile equating procedure was used to link the observed score distributions and produce the concordance tables. Equipercentile linking is a method that aligns the percentile ranks across the score distributions for the two tests. It has the benefit of being symmetrical, since the same percentile rank on the two tests represents the same position in the underlying ability distribution (Kolen & Brennan, 2014; Pommerich et al., 2004). Equipercentile linking is also intuitively interpretable, making it straightforward for testing organisations to convey the meaning of the score relationship to score users. Equipercentile linking is particularly suited to score use for high stakes selection decisions, where it is important that comparable selection criteria can be identified for both tests and that a similar proportion of the same group of test takers would be able meet them using either test.

The equipercentile linking analysis was conducted in R Studio using the equate package (Albano, 2016; R Core Team, 2023). Loglinear presmoothing was implemented jointly on both score distributions, as is advisable for equipercentile linking from a single group design (Albano, 2011; Dorans et al., 2010). The final concordance tables are based the equated IELTS Academic scores for each PTE Academic score point. PTE Academic is taken as the reference scale because it is more granular, allowing for more precise thresholds to be identified. This means that for each PTE Academic score point, an equipercentile equated IELTS score was identified from the pre-smoothed score distributions and rounded to the nearest 0.5 in line with the IELTS scale. The lowest PTE score that relates to an IELTS band score is considered the threshold score for that band. A separate linking model was created for overall score and each of the skill scores.

Concordance findings

Concordance summary table

The score concordance tables for overall score and each of the four skill scores are summarized in Table 8.

Table 8: *Concordance table for PTE Academic and IELTS Academic*

	PTE scores for each IELTS score band				
IELTS Score	**Overall**	**Listening**	**Reading**	**Speaking**	**Writing**
4.5	**24-30**	26-32	29-35	14-23	17-28
5.0	**31-38**	33-39	36-41	24-38	29-40
5.5	**39-46**	40-46	42-47	39-53	41-50
6.0	**47-54**	47-52	48-53	54-65	51-59
6.5	**55-62**	53-57	54-58	66-75	60-68
7.0	**63-70**	58-63	59-64	76-82	69-76
7.5	**71-78**	64-68	65-69	83-87	77-84
8.0	**79-85**	69-74	70-74	88-89	85-89
8.5	**86-89**	75-80	75-80	90	90
9.0	**90**	81-90	81-90	-	-

Concordance Interpretation

To support the interpretation of concordance reliability at different scores points, a table is provided below for each score type, including additional information. Tables 9.1-9.5 show the concordance sample size for each IELTS band along with standard errors of equating (SEE[1]). The sample size is useful in understanding the amount of data underpinning the concordance at different levels. SEE values are helpful in understanding the amount of variability in the concordance across the score range. Concordance scores with higher SEE values indicate that the data underlying the concordance were more variable at this level. Concordance scores with lower SEE values indicate that data were more consistent, and the concordance values are more reliable at these levels.

1 Bootstrap SEE values were calculated over 1,000 replications, drawing new samples with replacement while maintaining sample sizes. The SEE is the standard deviation of equated scores across all replications.

The sample sizes are most robust across all score types in the range of IELTS 5.5 to 8, and the SEE values are below 1 point across all scores in the ranges of IELTS 6 to 9. Concordance scores within these ranges are considered the most robust and reliable. It is also important to consider how the concordance tables function in relation to populations of interest, such gender, language, and nationality subgroups. Population invariance was investigated and found to be reasonable across the score ranges important for visa classification. See *Appendix C – Population invariance analysis* for full details.

Table 9.1: *Overall score concordance, sample size, and SEE*

IELTS Overall	PTE Overall	Sample Size	SEE
4.5	24-30	10	1.72
5.0	31-38	66	1.14
5.5	39-46	163	0.79
6.0	47-54	238	0.63
6.5	55-62	304	0.56
7.0	63-70	316	0.52
7.5	71-78	258	0.56
8.0	79-85	128	0.64
8.5	86-89	28	0.52
9.0	90	1	0.16

Table 9.2: *Listening score concordance, sample size, and SEE*

IELTS Listening	PTE Listening	Sample Size	SEE
4.5	26-32	29	1.62
5.0	33-39	91	1.09
5.5	40-46	144	0.79
6.0	47-52	189	0.69
6.5	53-57	174	0.66
7.0	58-63	147	0.63
7.5	64-68	236	0.60
8.0	69-74	166	0.61
8.5	75-80	235	0.67
9.0	81-90	98	0.59

Table 9.3: *Reading score concordance, sample size, and SEE*

IELTS Reading	PTE Reading	Sample Size	SEE
4.5	29-35	55	1.04
5.0	36-41	106	0.78
5.5	42-47	158	0.64
6.0	48-53	211	0.60
6.5	54-58	232	0.59
7.0	59-64	218	0.58
7.5	65-69	177	0.60
8.0	70-74	150	0.65
8.5	75-80	123	0.73
9.0	81-90	66	0.67

Table 9.4: *Speaking score concordance, sample size, and SEE*

IELTS Speaking	PTE Speaking	Sample Size	SEE
4.5	14-23	28	2.11
5.0	24-38	79	1.88
5.5	39-53	205	1.10
6.0	54-65	384	0.79
6.5	66-75	313	0.68
7.0	76-82	252	0.61
7.5	83-87	146	0.48
8.0	88-89	56	0.30
8.5	90	32	0.14
9.0	-	11	0.04

Table 9.5: *Writing score concordance, sample size, and SEE*

IELTS Writing	PTE Writing	Sample Size	SEE
4.5	17-28	12	3.04
5.0	29-40	45	1.33
5.5	41-50	186	0.63
6.0	51-59	543	0.40
6.5	60-68	441	0.37
7.0	69-76	212	0.50
7.5	77-84	52	0.71
8.0	85-89	15	0.84
8.5	90	2	0.25
9.0	-	0	0.00

Limitations of score concordance

Score concordance is a useful tool for institutions, governments, and other score users to support setting comparable requirements across different tests. While this concordance analysis has been based on a robust representative sample and carried out following rigorous good practice principles, score concordance is always an approximate relationship and should be interpreted with caution.

PTE Academic and IELTS Academic have been shown to have sufficiently similar constructs to support a score concordance, but they are ultimately distinct tests, with different structures, formats, and scoring criteria. The tests use different score scales with different levels of granularity, and so a one-to-one mapping of scores is always only approximate. The score scales may be used differently by each test, meaning that few test takers achieve scores in certain ranges of the scale, making concordance less precise in these regions. Score concordance is strongest in the regions of the scales where most test takers achieve scores. It is important to consider the sample size and Standard Error of Equating at the levels where score requirements are being set.

Concordance studies can be conducted using different methods, which may produce different results. Even the same well-designed study repeated at a different time with a different sample of test takers may show some degree of variation. This variation can be minimized by following good practice criteria, as this concordance study has, but it cannot be eliminated completely. Concordance studies are by their nature approximative.

Score requirements for university entrance, visa applications, or other high stakes decisions should be based on an understanding of the required performance standard, as well as test score concordance evidence. These requirements should be reviewed regularly and are subject to change over time. Pearson has and will continue to support governments and institutions in setting the right standards for their needs.

Conclusion

The purpose of this study has been to provide updated score concordance tables for the enhanced PTE Academic and the IELTS Academic tests to reflect the current relationship between their observed test scores. The concordance study followed in detail all the good practice principles for conducting concordance studies developed by Knoch & Fan (2024) ensuring robust score comparisons.

- An independent external construct comparison was carried out by the University of Bedfordshire Centre for Research in English Language Learning and Assessment (CRELLA). Their study concluded that the two tests were found to have overall comparable constructs.
- A large representative sample of 1,522 participants completed both tests within a 90-day period and in a reasonably counterbalanced testing order. The average time between tests was 18 days.
- Concordance study participants took both tests in controlled testing conditions in secure test centers and provided official score reports.
- The participants stated reasons for taking language tests reflect the range of reasons for testing found in the global testing population.
- The test taker survey indicated that participants were highly familiar with both tests and engaged in preparation activities to familiarize themselves with both tests.
- Descriptive statistics, full methodological details, statistical methods and procedures and correlation coefficients are described and shown in the report.
- A population invariance study was conducted, showing reasonable population invariance across score ranges important in high stakes decisions.
- Concordance results for both overall test and separate skill scores are shown in the report.
- The sample size and Standard Errors of Equating (SEE) are provided to support interpretation of the concordance tables.

In summary, this study has exceeded previous concordance studies for PTE Academic and IELTS Academic in terms of sample size and the granularity of concordance reporting at skill level. The rigor of the study along with the expanded information provided should serve to support institutions, governments, and other score users in making sound decisions about language proficiency in relation to academic, professional, or economic migration requirements. Alignments can and do change over time, whether from revisions to test design or evolution of the social context surrounding the tests. Therefore, updated concordance studies play a vital role in producing up to date and reliable information to guide the valid and appropriate use of test scores.

References

Albano, A. (2011). Statistical equating methods. *The R Project for Statistical Computing.*

Albano, A. (2016). equate: An R Package for Observed-Score Linking and Equating. *Journal of Statistical Software*, 74(8).

Chan, S., & Jones, J. (2025). *Establishing score concordance between the updated PTE Academic and IELTS Academic – Background: Construct Comparison.* Centre for Research in English Language Learning and Assessment (CRELLA).

Clesham, R., & Hughes, S. (2020). 2020 *Concordance Report for PTE Academic* and IELTS Academic. Pearson.

Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching assessment (CEFR). A manual*. Language Policy Division.

Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). Principles and Practices of Test Score Equating. *ETS Research Report Series*, 2010(2), i–41.

Feldt, L., & Brennan, R. (1989). Reliability. In R. Linn (Ed.), *Educational Measurement* (3rd ed.). American Council on Education.

IELTS. (2024). *Test Statistics*. IELTS

IELTS Test Performance Data 2023-2024. (2024).

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

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer-Verlag.

Linacre, J. (2024). Reliability and separation of measures. In A User's Guide to *Winsteps Rasch-Model Computer Programs* (v5.7.4).

Pommerich, M., Hanson, B. A., Harris, D. J., & Sconing, J. A. (2004). Issues in Conducting Linkages between Distinct Tests. *Applied Psychological Measurement*, 28(4), 247–273.

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 2023.12.0+369) [Computer software]. R Foundation for Statistical Computing.

Zheng, & De Jong. (2011). *Establishing Construct and Concurrent Validity of Pearson Test of English Academic* [Research Note]. Pearson.

Acknowledgements

We are grateful to the panel of internationally recognised experts who have supported the research described in this document, including our external reviewers, Dr. Jason Fan (Deputy Director and Principal Research Fellow at the Language Testing Research Centre (LTRC), University of Melbourne), Dr. Anthony Green (Director of the Centre for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire), and Dr. Jon Twing (Honorary Research Fellow at the University of Oxford in England, and at the University of Sydney in Australia), who reviewed early drafts of this report and provided invaluable methodological guidance; and to the co-authors of Establishing score concordance between the enhanced PTE Academic and IELTS Academic – Background: Construct Comparison, Dr. Sathena Chan (Associate Professor, CRELLA University of Bedfordshire) and Dr. Johnathan Jones (Lecturer, CRELLA University of Bedfordshire), who carried out an independent investigation into construct comparability, which underpins the enhanced PTE Academic and IELTS Academic concordance.

Appendices

Appendix A – Test order effect analysis

The concordance study was successful in collecting an equally counterbalanced sample that exceeds 1,000 test takers (n=1,109), meaning that the same number of people sat IELTS first (n=551) as sat PTE Academic first (n=551), and a small number sat both tests on the same day (n=7). This sample was supplemented with additional data to increase the robustness of score comparisons, and this data had an IELTS-first testing order (n=413). When combined, the full sample (n=1,522) has more representation of the IELTS-first testing order (n=964) than PTE-first testing order (n=551).

While the full sample remains reasonably counterbalanced, an analysis was conducted to evaluate whether the additional IELTS-first data should be used in the concordance analysis. This analysis included comparing differences in concordance tables calculated on the equally counterbalanced sample and the sample with additional data, as well as an analysis of potential order effects in the scoring data.

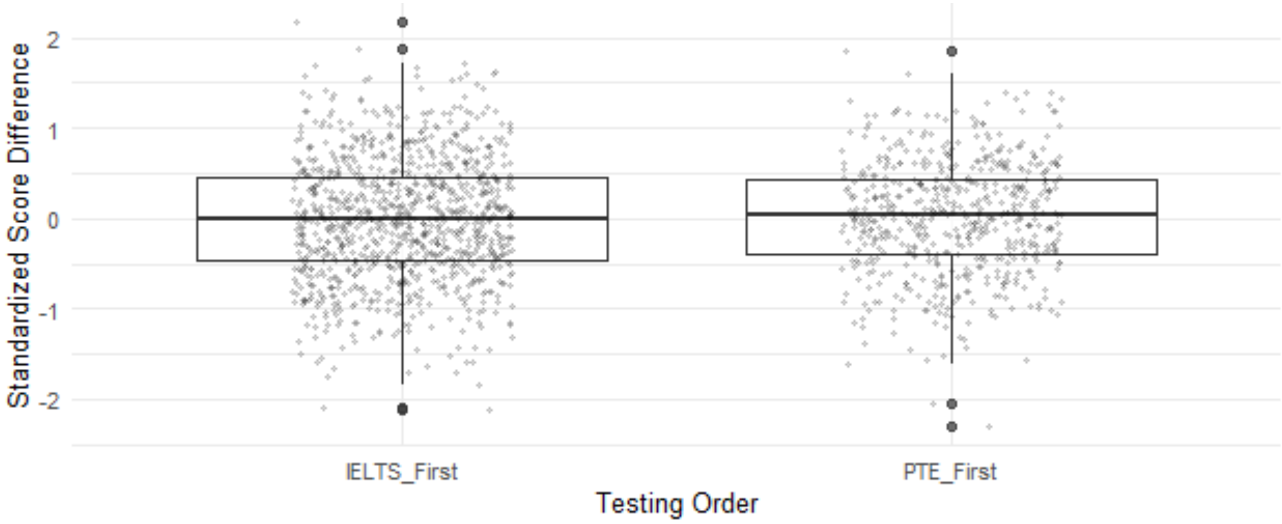
Table 10 shows the concordance table based on all data compared to the concordance table based on only the equally counterbalanced sample. To create the equally counterbalanced concordance values, all PTE-first and same-day test takers were included, along with a random sample of IELTS-first test takers of the same number as PTE-first test takers.

Table 10: *Comparing concordance values based on full sample (n=1,522) and equally counterbalanced subsample (n=1,109)*

	Overall		Listening		Reading		Speaking		Writing	
IELTS	With Additional data	Counter-balanced	With Additional data	Counter-balanced	With Additional data	Counter-balanced	With Additional data	Counter-balanced	With Additional data	Counter-balanced
4.5	24	24	26	26	29	30	14	14	17	15
5.0	31	31	33	33	36	36	24	23	29	28
5.5	39	39	40	40	42	42	39	38	41	41
6.0	47	47	47	47	48	48	54	53	51	52
6.5	55	55	53	53	54	54	66	66	60	60
7.0	63	63	58	58	59	59	76	76	69	69
7.5	71	71	64	64	65	65	83	83	77	77
8.0	79	79	69	69	70	70	88	88	85	84
8.5	86	86	75	75	75	75	90	90	90	90
9.0	90	90	81	81	81	81	-	-	-	-

The resulting concordance table for the equally counterbalanced sample is nearly identical to the sample with additional data. The concordance values for overall score are exactly the same, with a small number of skill score thresholds shifting by 1-2 points at levels where the sample size was most impacted by excluding the additional data. These shifts are minimal and are best explained by the loss in precision of excluding 413 test takers, rather than a potential order effect in the data. To evaluate potential testing order effects on scoring, PTE and IELTS scores were standardized and evaluated with an independent samples t-test and one-way ANOVA. Standardized score differences between the first and second test were found to be normally distributed for both PTE-first and IELTS-first testing orders using the Shapiro-Wilk test ($W = 0.999$, $p = 0.726$). Figure 9 shows the distribution of standardized score differences by group.

Figure 9: *Standardized score differences by testing order*



An independent samples t-test was selected to evaluate the order effect. Though there are paired testing observations for each test taker, we are most interested in understanding whether the difference between these observations varies significantly depending on test order group. Test order groups are completely independent samples (i.e. it is only possible for a test taker to belong to one test order group). The independent samples t-test was conducted on standardized score differences for the PTE-first group ($M = 0.012$, $SD = 0.626$, $n=551$) and the IELTS-first group ($M = -0.009$, $SD = 0.680$, $n=964$). The t-test found no significant difference between testing order groups ($t(1223.4) = -0.608$, $p = 0.543$, 95% CI [-0.089, 0.047]), suggesting that test order did not systematically influence performance. The one-way ANOVA supported the findings of the t-test, with no significant difference found ($F(1, 1513) = 0.354$, $p = 0.552$). The effect size $\eta^2 = 0.00023$ indicates that only 0.023% of the observed variance was attributable to test order.

Ultimately, these findings demonstrate that there is negligible difference between a concordance calculated on the equally counterbalanced sample and a concordance calculated with additional data. Further, there is no evidence to suggest that these minimal differences are attributable to order effect, as there is no significant order effect evident in the scoring data. Therefore, the additional data is considered beneficial to the concordance analysis and the full sample of 1,522 participants was retained.

Appendix B – Enhanced PTE Academic Conditional Standard Error of Measurement

The Conditional Standard Error of Measurement (CSEM) table provided below is based on all field testing and concordance study data for the enhanced PTE Academic test.

Table 11: *Enhanced PTE Academic CSEM table (n = 4,327)*

CEFR Level	Overall	Listening	Reading	Speaking	Writing
A2	3.1	4.4	4.8	6.2	4.9
B1	2.8	3.8	4.8	5.5	4.1
B2	3.2	4.5	5.8	5.6	4.8
C1	4.0	5.6	7.3	6.6	6.3
C2	4.5	6.5	8.4	7.8	7.0
All Categories	**3.1**	**4.2**	**5.2**	**5.7**	**4.6**

Appendix C – Population invariance analysis

Population invariance is a necessary and important aspect in understanding the relevance of concordance data to different groups of test takers. However, it requires careful interpretation, as it relies on dividing the concordance sample into subgroups, which results in less reliable concordance calculations. This analysis focuses on three test taker characteristics and defines the subgroups within each category to maximize sample size.

- **Gender** –Two largest gender subgroups, Male and Female (Tables 12-14)
- **Language** – Largest language subgroup (Mandarin-Chinese) in comparison to all other languages together (Tables 15-17)
- **Nationality** – Two largest nationality subgroups, India and China (Tables 18-20)

The following analysis identifies differences in concordance scores calculated on subgroup populations at IELTS 4.5-8.5, and compares the underlying subgroup score distributions. On each concordance table, (Tables 14, 17 and 20) cells are in highlighted in grey where sample sizes are less than 20, and cells are highlighted in blue where the difference between subgroup concordance scores is greater than 3.

For all three characteristics – gender, language, and nationality – no meaningful concordance variance was found for Overall scores.

Some concordance variation was found for specific skill scores at specific IELTS levels. Of the 108 concordance skill score comparisons in Tables 14, 17 and 20, 11 of them show difference between subgroups of more than 3 points. This an expected finding, given that skill subscores by definition are based on only a subset of test questions, making them less reliable than overall scores. This effect is then magnified when calculating concordance between subscores using smaller subgroup samples at each level.

Overall, the analysis shows similar score distribution patterns for subgroups of test takers across both tests, and the resulting concordance tables show reasonable population invariance across score ranges important in high stakes decisions.

Gender

Table 12: *Mean and standard deviation of concordance sample PTE scores by gender subgroup*

		Overall		Listening		Reading		Speaking		Writing	
Gender	n	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Female	800	6.69	0.89	7.06	1.29	6.76	1.23	6.42	0.89	6.26	0.64
Male	722	6.62	0.88	7.00	1.22	6.65	1.21	6.35	0.91	6.20	0.67

Table 13: *Mean and standard deviation of concordance sample IELTS scores by gender subgroup*

		Overall		Listening		Reading		Speaking		Writing	
Gender	n	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Female	800	60.92	13.51	60.55	14.90	58.62	13.10	64.7	18.35	59.14	11.17
Male	722	60.61	13.85	60.67	14.92	57.31	14.47	64.65	17.64	58.79	11.50

Figure 10: *Score distribution density plots for concordance sample IELTS and PTE scores by gender subgroup*

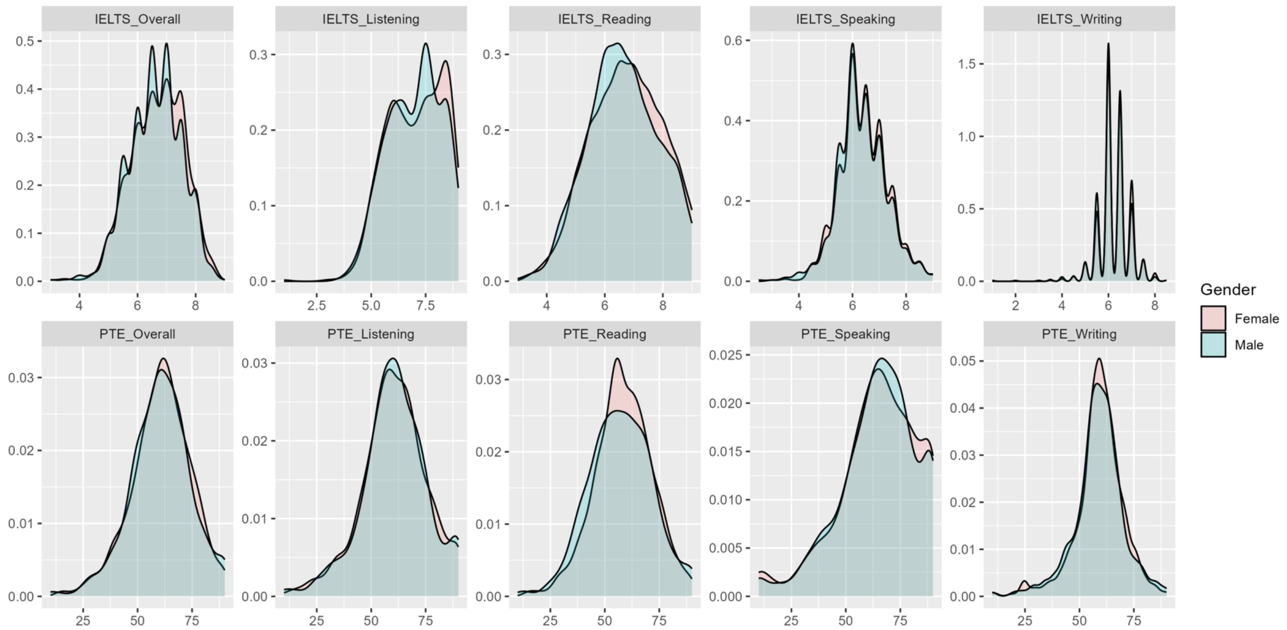


Table 14: *Concordance values calculated by gender subgroup*

	Overall		Listening		Reading		Speaking		Writing	
IELTS	F	M	F	M	F	M	F	M	F	M
4.5	24	24	27	24	31	28	12	16	14	21
5.0	31	31	34	32	37	35	20	27	27	31
5.5	39	39	40	40	43	41	36	41	41	42
6.0	47	47	47	46	49	47	52	55	51	51
6.5	55	55	52	53	54	53	65	66	60	60
7.0	62	63	58	59	59	59	76	76	68	69
7.5	70	72	63	65	64	65	83	83	76	78
8.0	78	80	68	70	69	71	88	88	84	86
8.5	85	86	74	75	74	76	90	90	90	90

Language

Table 15: *Mean and standard deviation of concordance sample PTE scores by language subgroup*

		Overall		Listening		Reading		Speaking		Writing	
Language	n	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mandarin	637	6.71	0.92	7.07	1.30	6.93	1.22	6.32	0.91	6.23	0.71
Other language	885	6.62	0.86	7.00	1.23	6.55	1.20	6.44	0.89	6.23	0.61

Table 16: *Mean and standard deviation of concordance sample IELTS scores by language subgroup*

		Overall		Listening		Reading		Speaking		Writing	
Language	n	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mandarin	637	60.54	14.25	59.58	15.88	58.87	13.89	64.04	18.28	58.6	12.54
Other language	885	60.94	13.24	61.34	14.12	57.38	13.66	65.14	17.81	59.24	10.36

Figure 11: *Score distribution density plots for concordance sample IELTS and PTE scores by language subgroup*

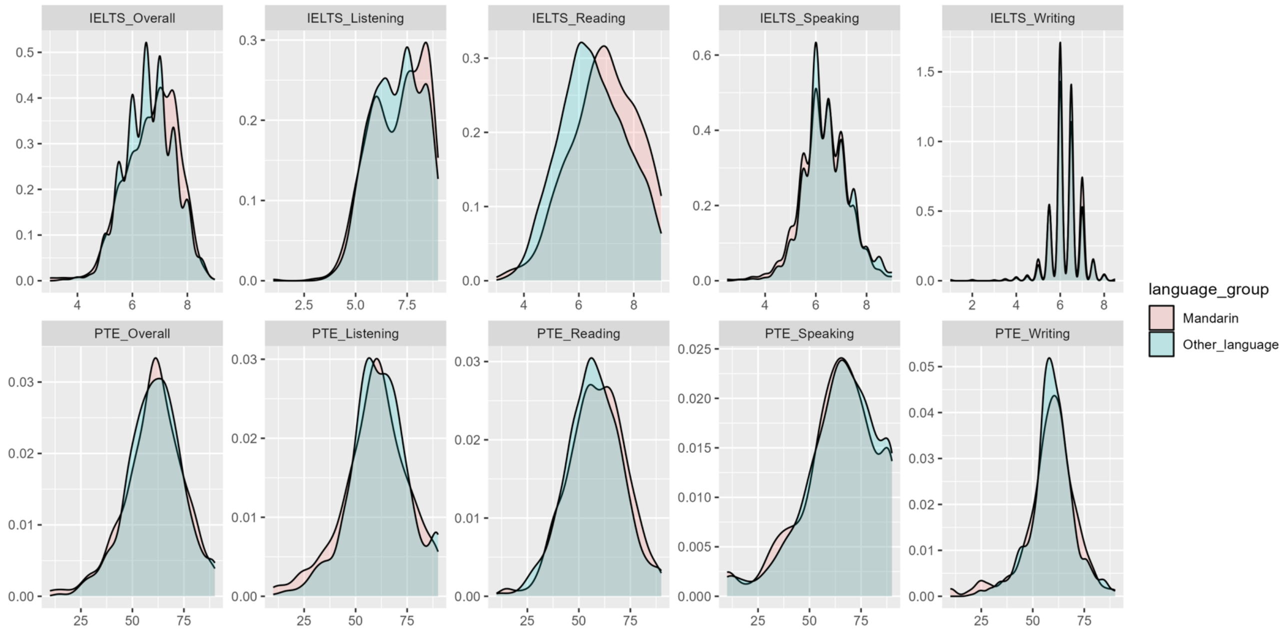


Table 17: *Concordance values calculated by language subgroup*

	Overall		Listening		Reading		Speaking		Writing	
IELTS	Mandarin	Other	Mandarin	Other	Mandarin	Other	Mandarin	Other	Mandarin	Other
4.5	22	24	22	30	29	29	17	12	13	24
5.0	29	32	30	36	35	36	27	21	24	33
5.5	37	40	38	42	40	43	41	37	40	42
6.0	46	48	45	48	46	49	54	53	51	51
6.5	54	56	51	53	52	55	66	66	60	60
7.0	62	64	57	59	57	61	76	76	68	69
7.5	70	72	63	64	63	66	83	83	76	77
8.0	78	79	68	70	68	71	88	87	84	86
8.5	85	86	73	76	74	76	90	90	90	90

Nationality

Table 18: *Mean and standard deviation of concordance sample PTE scores by nationality subgroup*

		Overall		Listening		Reading		Speaking		Writing	
Nationality	n	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
China	432	6.66	0.89	6.95	1.28	6.98	1.21	6.2	0.86	6.22	0.62
India	332	6.42	0.79	6.88	1.14	6.21	1.12	6.22	0.78	6.12	0.52

Table 19: *Mean and standard deviation of concordance sample IELTS scores by nationality subgroup*

		Overall		Listening		Reading		Speaking		Writing	
Nationality	n	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
China	432	59.45	14.25	58	15.84	58.48	13.62	62.78	18.67	57.93	12.37
India	332	56.65	13.69	57.89	14.25	51.98	13.87	60.98	19.34	55.8	10.67

Figure 12: *Score distribution density plots for concordance sample IELTS and PTE scores by nationality subgroup*

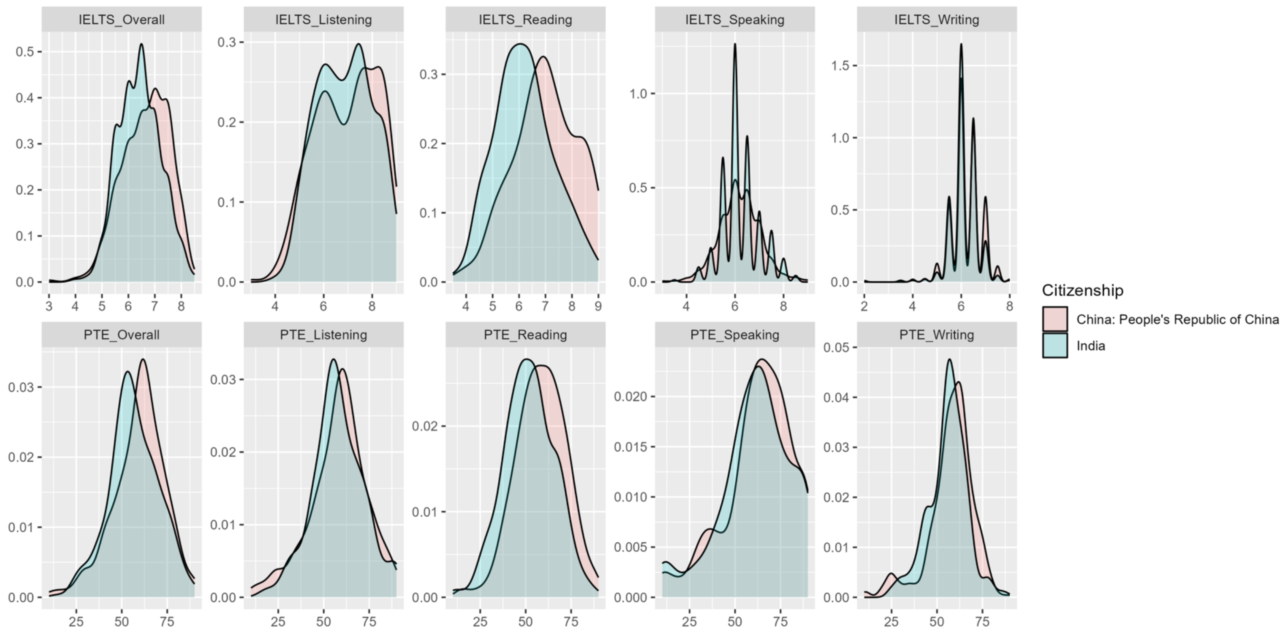


Table 20: *Concordance values calculated by nationality subgroup*

	Overall		Listening		Reading		Speaking		Writing	
IELTS	China	India	China	India	China	India	China	India	China	India
4.5	18	18	19	24	26	25	13	10	11	17
5.0	26	27	28	31	33	33	23	14	16	27
5.5	36	36	37	38	40	41	40	31	37	38
6.0	45	45	45	44	46	48	56	52	51	48
6.5	54	54	52	51	51	54	68	67	60	59
7.0	62	63	57	57	57	60	77	77	69	70
7.5	70	72	63	63	62	65	84	84	77	80
8.0	77	80	68	69	67	70	89	88	85	89
8.5	84	87	73	75	72	75	90	90	90	90

Pearson

Secure. Accurate. Trusted.

pearsonpte.com/research