

2020 Concordance Report

PTE Academic and IELTS Academic

November 2020

Dr. Rose Clesham

Sarah R. Hughes

Contents

Abstract.....	3
Introduction	3
Alignment in context	3
Purpose of this study	4
Rationale for Linking PTE Academic and IELTS Academic.....	4
Methodology	6
Study design overview	6
Sample description.....	7
Analysis	8
Concordance Table.....	11
Interpretation of linking.....	12
Conclusion.....	12
References.....	13
Appendix 1 – Comparison of PTE Academic and IELTS Academic tests	15

Abstract

Alignment studies are important in supporting the interpretation of test scores and standards. This paper reports on the results of a particular type of alignment, score concordance between PTE Academic and IELTS Academic tests. Scores for these tests carry significant currency in terms of academic, professional and economic migration entry requirements and a concordance table supports the interpretation of the relationship between these scores.

The original concordance between PTE Academic and IELTS Academic test scores is based on a linking study conducted during the initial field testing of PTE Academic (Zheng & De Jong, 2011). The purpose of the current research study is to update the original concordance table based on the testing data and research accumulated in the decade since the launch of PTE Academic.

Introduction

Alignment in context

At the core of the language teaching, learning, and assessment, there are three key alignments:

1. The alignment of learning objectives to curriculum or content standards
2. The alignment of content standards to performance standards (test outcomes)
3. The alignment of different performance standards and measurement scales

This paper addresses the third type of alignment, the alignment of measurement scales for two different tests. However, it is important to understand this type of alignment in the context of the others.

The first type of alignment, the alignment of language learning objectives to a recognised schema or framework, for example the Common European Framework of Reference (CEFR) (Council of Europe, 2001), is essential in order to support teaching and learning programmes and to indicate general levels of progression or attainment. Most language content standards are not designed to be treated empirically, as they describe proficiency rather than performance. As such there are no definitive empirical alignments between content standards, just best fits.

The second type of alignment, that of performance standards to content standards, defines the relationship between test scores and recognised standards of proficiency. Guidance and methodologies are provided in the *Manual for relating Language Examinations to the Common European Framework of Reference for Languages* (Council of Europe, 2009), following the stages of familiarisation, specification, standardisation and validation of both frameworks and tests. Having said this, it should be remembered that different methodologies can lead to different alignment outcomes and human judgemental exercises can be influenced by unconscious bias and heuristics (Eckes, 2012).

The third type of alignment is the focus of this paper: aligning the performance standards of different tests by conducting a linking study between their score reporting scales. In many ways linking studies should be the easiest alignment exercise because they simply compare score performance data. The simple question is what are the comparable score points on two distinct tests? Although the concept of implementing a linking study sounds simple, the detail is important. What is the linking study design, how is the sample selected, how representative is the sample across the measurement range of the tests, how much time has elapsed between test administrations, how is test data collected and what is the rationale for the selected linking

methodology? These questions have significant bearing on the outcome of any linking study. Any performance standards comparative analyses work is further complicated as the tests themselves are invariably somewhat different in terms of the test constructs, item types, scoring rubrics, marking methods and standard setting procedures.

It is the responsibility of testing organisations to investigate and update their reported alignments to support the interpretation and use of test scores. In the case of performance standard alignment, the resulting score concordance can carry significant currency in terms of academic, professional and economic migration entry requirements.

Purpose of this study

In the context of international high stakes English language tests, test scores from different tests are often used for the same purpose. For example, university admissions or immigration visa applications may require applicants to demonstrate a specified level of English language proficiency. These institutions may recognise a number of different English language tests as appropriate means to demonstrate proficiency. Ideally, institutions would specify required scores for each test in relation to the desired performance standard on that test. However, it may also be useful to understand how the score scales on the two different tests relate to each other. If the tests are shown to be reasonably similar, their score scales can be “linked”, and a concordance table can be produced to show comparable score points along their measurement scales. It is important to note the difference between “linking” and “equating”. Because these different tests may not be identical in their task and response types, assessed constructs, content coverage, timing, or scoring methodology, it is not possible to “equate” their score scales precisely. The differences in the tests themselves and each test’s individual measurement error mean that the linking relationship will always be approximate. With this in mind, score concordance tables can be a useful tool in supporting the interpretation and use of scores in international high stakes contexts where multiple testing organisations operate.

The original concordance between PTE Academic and IELTS Academic test scores is based on a linking study conducted during the initial field testing of PTE Academic (Zheng & De Jong, 2011). The purpose of the current research study is to update the original concordance table based on the testing data and research accumulated in the decade since the launch of PTE Academic.

Rationale for Linking PTE Academic and IELTS Academic

Scores from two different tests can only be linked if the tests share similar characteristics and statistical properties (Kolen & Brennan, 2014). Prior research has established the appropriateness of linking PTE Academic and IELTS scores based on their similar test purposes, test score uses, assessed constructs, task types, and statistical reliabilities (De Jong & Benigno, 2017). These conditions, which are summarised below and in Appendix I, remain unchanged and it continues to be both appropriate and necessary to provide a score concordance table between PTE Academic and IELTS Academic tests.

Test purpose and assessed constructs

Both tests claim their purpose is to assess the English language proficiency required for international work, study, or immigration, and both tests are used in high stakes decisions in these contexts. As claimed by the test providers (Taylor, 2004; Zheng & De Jong, 2011), the constructs of both tests are embedded within the descriptors of the CEFR, which provides a

common frame of reference for different assessment organisations to describe their approach to the assessing the domain of English language.

The Global Scale of English (GSE) (De Jong, Mayor, & Hayes, 2016) was first applied as the reporting scale for the PTE Academic test. The test was designed to align to the CEFR and was developed using the procedures recommended in the *Manual for relating Language Examinations to the Common European Framework of Reference for Languages* (Council of Europe, 2009). As a result of aligning the design of PTE Academic to the CEFR, the GSE reporting scale is a linear transformation of the logit scale underlying the CEFR descriptors developed by North (2000) and, ranging from 10 to 90, provides a more granular measurement of performance. Extensive testing has been undertaken to ensure that the relationship between the GSE and the CEFR is supported by statistical data. More information can be found at pearsonpte.com/research (De Jong & Zheng, 2016; Pearson, 2010).

IELTS results are reported on a 9-band scale (including half bands), ranging from Non language (1) to Expert language users (9). The IELTS test was developed before the development of the CEFR and has been subsequently mapped to the CEFR. While IELTS asserts that there is not a one-to-one correspondence between IELTS scores and CEFR levels, Cambridge ESOL has conducted a number of research projects since the late 1990's to explore how IELTS band scores align with the CEFR levels (Lim, Geranpayeh, Khalifa, & Buckendahl, 2013). Taylor (2004) summarises a number of these studies while noting, "As we grow in our understanding of the relationship between IELTS and the CEFR levels, so the frame of reference may need to be revised accordingly."

The individual alignments of each test to the CEFR are not the focus of this study, however, it is important for any linking study to establish that the tests being linked intend to measure similar constructs. As both PTE Academic and IELTS Academic have described similar performance standards in a common language of English proficiency, this supports the appropriateness of linking their score scales empirically.

Test design and administration

PTE Academic is a computer-based international English language test. It assesses test takers' English language competency in listening, reading, speaking and writing. The test uses 20 item types reflecting different modes of language use and response formats through a combination of single skill and integrated skills tasks. The maximum duration of the test is three hours and is administered entirely on computer in secure test centres using Pearson's state-of-the-art security measures (Lopes, 2010).

IELTS is an international language test, jointly owned by the British Council, IDP: IELTS Australia and Cambridge Assessment English. IELTS also assesses all four communicative skills. The listening, reading and writing components are completed in a secure examination setting on the same day, and can be delivered either on paper or computer. The speaking component is assessed separately through a face-to-face interview up to a week before or after the written test. The reading and listening sections consist of closed question types. Writing is assessed through two task types (information synthesis and essay based) and speaking is assessed through three task types (everyday questions, a monologue and a discussion). The total test time is 2 hours and 45 minutes.

Both tests have a similar test duration and structure, and they assess of four core communicative skills, though the tests differ in their approach to the assessment of those skills. IELTS item types assess individual skills, whereas PTE Academic uses a mixture of item types assessing individual and integrated skills. The speaking portions of the tests differ significantly. PTE Academic is entirely computer-based, whereas IELTS uses a live interview format. PTE Academic and IELTS Academic are broadly similar in design, and the differences between the tests can provide context for interpreting the results of a linking study.

Standard Error of Measurement (SEM)

The combined effect of the Standard Error of Measurement (SEM) for both tests will impact the precision of any linking study. The SEM for PTE Academic is 2.3 GSE points and the SEM for IELTS is 0.23 of an IELTS band (Pearson, 2014). Both tests are used in high stakes decisions and are sufficiently reliable to support linking their score scales.

Methodology

Study design overview

The linking study comprised two phases of data collection. Each phase used a single group design and was generally counterbalanced for testing order, with approximately half of the test takers completing PTE Academic first, and the other half completing IELTS first.

In Phase 1, Pearson collected self-reported scores in exchange for a small monetary incentive from test takers who had already completed both tests of their own accord. The score collection method used in Phase 1 is commonly employed in linking studies because it has the benefit of collecting scores from intrinsically motivated test takers. However, it can be challenging to collect scores from across the ability spectrum with this method alone. Phase 2 was implemented to ensure a robust sample size and representative ability spectrum.

In Phase 2, test takers were selected from a sample of survey respondents interested in test preparation for PTE Academic and IELTS Academic. Test takers received free preparatory courses¹ to encourage familiarisation with both tests, they were provided with a voucher for a free PTE Academic test, and they were reimbursed for the cost of an IELTS test. Test takers were assigned a testing order to ensure a counter balanced design and instructed to complete both tests within one month. Test takers agreed to have their PTE Academic score held until the completion of the testing programme to ensure that the results would not influence their performance on IELTS. This was necessary for PTE Academic, as most scores are typically returned within 2 days. IELTS scores are typically returned within approximately 2 weeks, and many test takers naturally completed their PTE Academic test within this time.

PTE Academic and IELTS Academic overall scores and communicative skill subscores were collected and analysed. Equipercentile equating was used to link the observed overall score distributions and produce a concordance table. Equipercentile equating is a method that aligns the percentile ranks across the score distributions for the two tests. It has the benefit of being symmetrical, meaning that regardless of which test is used as a reference point, the percentile ranks will be the same, and interpretable for test score users who need to identify comparable scores for different tests used in a selection process (Kolen & Brennan, 2014; Pommerich,

¹ Preparatory courses for both PTE Academic and IELTS Academic were provided by E2Language.

Hanson, Harris, & Sconing, 2004). Where test scores are used for selection decisions, 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.

Sample description

Data were collected over two phases in an effort to obtain a representative sample of the PTE Academic testing population and introduce measures to minimise bias where possible.

The data set includes pairs of PTE Academic and IELTS Academic overall scores for 562 unique test takers. In most cases, test takers also reported their subscores for listening, reading, speaking, and writing skills. While official PTE Academic score reports were available for all test takers, official IELTS score reports were provided by approximately half of the test takers. Table 1 shows the number of test takers, subscores, and score reports included in each phase of data collection.

Table 1. Data collected in each phase

Phase	Number of test takers who reported overall scores	Number of test takers who reported subscores	Number of test takers who provided official IELTS score reports
Phase 1	389	218	113
Phase 2	173	173	173
TOTAL	562	391	286

Demographic data were available for the full sample of 562 test takers. The data indicate that the sample is representative of the PTE Academic testing population in terms of age, primary language and country of citizenship.

The majority of the test takers were between the ages of 23 and 31, with an average age of 27.8. Test takers came from 59 countries and spoke 53 different primary languages. Figure 1 shows the full test taker age range, and Tables 2.1 and 2.2 show the five most common languages and nationalities in our sample. The most common languages represent about 49% of our sample and the most common nationalities represent about 66% of the sample.

Figure 1 – Test taker age

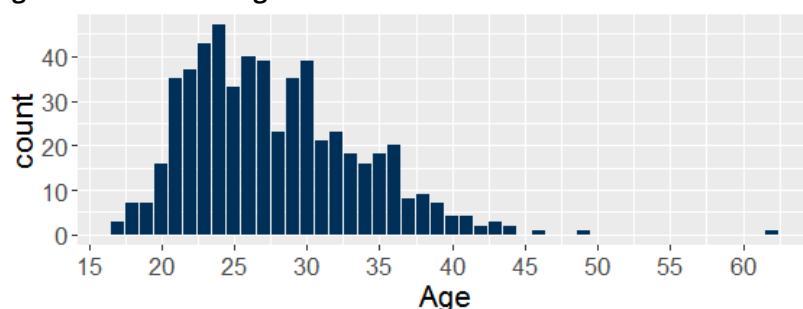


Table 2.1 – Top 5 languages spoken

Language	Count	%
Chinese-Mandarin	100	17.8%
English	58	10.3%
Urdu	45	8.0%
Hindi	40	7.1%
Telugu	31	5.5%

Table 2.2 – Top 5 countries of citizenship

Country	Count	%
India	159	28.3%
China	108	19.2%
Pakistan	46	8.2%
Philippines	31	5.5%
Nepal	28	5.0%

Test date information was available for 391 test takers. These data demonstrate a reasonably counterbalanced testing order, which minimises bias where the results from the first test may influence a test taker's performance on the second test. Table 3 shows balance of testing order where test date information was available.

Table 3 – Testing order

	Count	%
PTE Academic First	199	50.9%
IELTS First	192	49.1%

Test date information was also used to calculate the time between tests. In Phase 1, score reports were collected from test takers who had taken PTE Academic and IELTS of their own accord and on their own schedule. The time between tests in this phase was on average 72 days, with most tests being completed within 100 days. In Phase 2, Pearson recruited test takers for the specific purpose of completing both tests within a short period of time. The average time between tests in Phase 2 was only 19 days, with a maximum of 47 days. While the time between tests is longer in Phase 1, this is somewhat mitigated by the counterbalanced testing order, as the potential for score improvement during the time between tests was equal for both PTE Academic and IELTS.

The final sample of 562 test takers was compiled to be representative and to minimise bias while also maintaining a robust sample size to support the interpretation of the results.

Analysis

In order to equate two tests of similar purpose and design, the strength of the empirical relationship between the tests must also be established. The following analyses consider the representativeness of the sample and strength of the relationship between test scores.

The equipercentile analysis is based on the overall scores collected from 562 test takers. The test takers scored between 10 and 90 on PTE, with an average score of 63, and between 4 and 8.5 on IELTS, with an average score of 6.6.

Note that no test taker in our sample received the maximum overall score of 9 on IELTS. Similarly, publicly available IELTS data suggests that less than one percent of test takers receive an overall score of 9 in the live test setting ("IELTS Demographic Data," 2019). This score category is so infrequently accessed that there is insufficient data to equate this point to the PTE Academic scale.

Table 4 shows the cumulative frequency of scores in our sample in relation to score ranges based on the original concordance.

Table 4 – Cumulative frequency of overall scores

Score Range	IELTS (n)	PTE (n)
<30 PTE / 4.0 IELTS	2	8
30-35 PTE / 4.5 IELTS	8	21
36-41 PTE / 5.0 IELTS	23	50
42-49 PTE / 5.5 IELTS	79	118
50-57 PTE / 6.0 IELTS	181	207
58-64 PTE / 6.5 IELTS	322	294
65-72 PTE / 7.0 IELTS	427	403
73-78 PTE / 7.5 IELTS	507	469
79-82 PTE / 8.0 IELTS	554	502
83-85 PTE / 8.5 IELTS	562	520
>85 PTE / 9.0 IELTS	562	562

The majority of test takers in our sample fall approximately within PTE Academic 50 to 75 and IELTS 6 to 7. Official test scores for PTE Academic in 2019 show a similar distribution to the test scores in our sample, indicating that our sample is representative of the PTE Academic testing population. Figure 2 shows the 2019 PTE Academic scores as a reference point, along with the PTE Academic and IELTS Academic scores in our sample. Table 5 provides statistics to describe the overall score distributions for both tests in our sample and PTE Academic scores in 2019.

Figure 2 – Overall scores for PTE Academic in 2019 alongside overall scores in sample

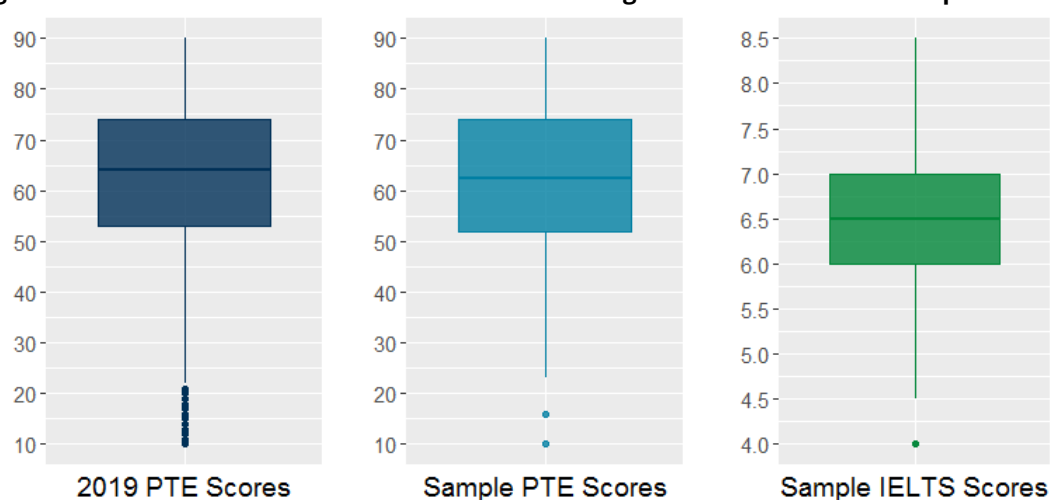
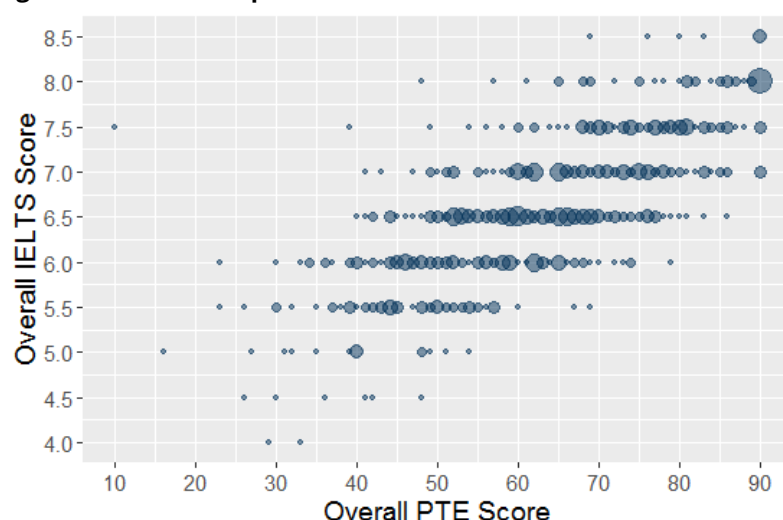


Table 5 – Overall score distribution information

	Mean	SD	Relative SD	Skew	Kurtosis	Min	Max
2019 PTE Academic	62.96	14.56	0.25	-0.32	2.70	10	90
Sample PTE Academic	62.57	15.30	0.24	-0.20	2.60	10	90
Sample IELTS Academic	6.63	0.83	0.12	-0.10	2.79	4	8.5

Figure 3 shows the relationship between overall scores on PTE Academic and IELTS Academic for each test taker in our sample. Each point represents one test taker and the size of the points has been scaled where test takers overlap. There was a strong correlation ($r=0.74$) between overall scores for PTE Academic and IELTS Academic.

Figure 3 – Relationship of PTE and IELTS Overall Scores



The relationship between the overall score distributions is sufficiently strong to support equipercentile equating. Although the overall scores are the basis of the concordance table, it is helpful to also consider the relationship of the subscores for listening, reading, speaking and writing. These communicative skill correlations are moderate, ranging from 0.42 to 0.68. Table 6 provides all skill-to-skill correlations. The same-skill correlations have been highlighted as the most relevant.

Table 6 – Communicative skill score correlations

	IELTS Listening	IELTS Reading	IELTS Speaking	IELTS Writing
PTE Listening	0.66	0.59	0.56	0.60
PTE Reading	0.68	0.68	0.53	0.56
PTE Speaking	0.49	0.38	0.42	0.41
PTE Writing	0.67	0.67	0.53	0.60

The correlations between communicative skill subscores are lower than the correlation between overall scores, and this is to be expected as the majority of PTE Academic items address integrated skills, whereas IELTS items address single skills. The score concordance between PTE Academic and IELTS Academic should be interpreted in this context for communicative skill subscores.

Concordance Table

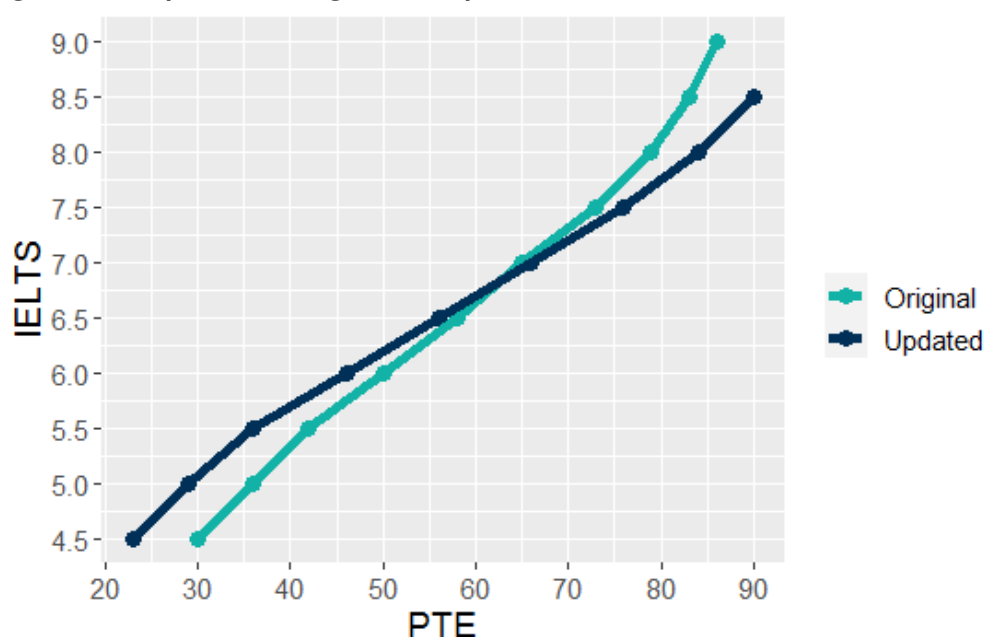
The equipercentile equating analysis was conducted in R Studio using the *equate* package (Albano, 2016). Loglinear presmoothing was implemented jointly on both score distributions, as is advisable for equipercentile equating from a single group design (Dorans, Moses, & Eignor, 2010).

For each PTE Academic score point, a percentile rank was calculated. The IELTS score of the same percentile rank was identified and rounded to the nearest 0.5 in line with the IELTS scale. The results of the equipercentile equating are shown in Table 7 and Figure 4 alongside the originally estimated concordance values (Zheng & De Jong, 2011).

Table 7 – Updated concordance table for PTE Academic and IELTS Academic

PTE (original)	PTE (updated)	IELTS
30	23	4.5
36	29	5.0
42	36	5.5
50	46	6.0
58	56	6.5
65	66	7.0
73	76	7.5
79	84	8.0
83	89	8.5

Figure 4 – Comparison of Original and Updated Concordance

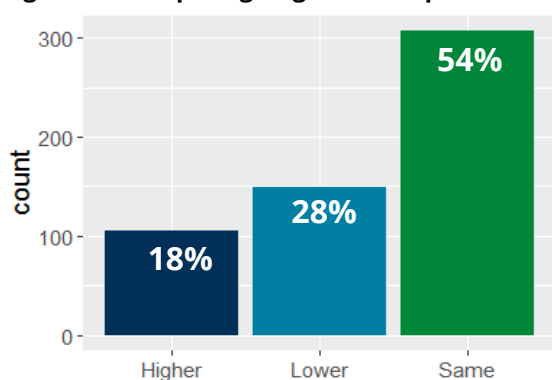


Interpretation of linking

The updated concordance table shows that the estimates for concordant scores have been adjusted up or down at different points on the scale. Above PTE Academic 65, the updated estimated concordant IELTS Academic scores are the same or lower. Below PTE Academic 65, the updated estimated concordant IELTS Academic scores are the same or higher. In the mid-range of the scale where most test takers fall, the updated estimates for concordant scores show minimal differences from the original estimates. The adjustments tend to grow larger toward the extremes of the scales.

To understand how this would impact individual test takers, we can compare the original and updated estimated concordant IELTS scores for the 562 test takers in our sample based on their PTE Academic score. Because most test takers obtain scores in the middle ranges, for most of the test takers in our sample (54%), their estimated concordant IELTS scores would not change under the updated table. For about 18%, the estimated concordant IELTS scores would be higher and for about 28% they would be lower, as shown on Figure 5.

Figure 5 – Comparing original and updated concordant IELTS scores in our sample



To understand if this change better reflects reality, we can compare the updated estimated concordant IELTS scores based on each test taker's PTE Academic score with the actual IELTS scores these test takers received. The average absolute difference between the observed IELTS scores and the updated estimated concordant IELTS scores for test takers in our sample is 0.45 of an IELTS band. Using the original concordance, this difference is larger, 0.60 of an IELTS band. The updated estimated concordant IELTS scores are closer to the observed IELTS scores for individuals in our sample. While concordance tables will never achieve absolute precision for each individual test taker, this updated concordance more closely aligns comparable scores across the testing population.

Conclusion

The purpose of this linking study has been to update the estimated concordance test score values between the PTE Academic and IELTS Academic tests. The tests themselves continue with differing test designs and test modes, however, their currency in terms of entry requirements for academic, professional or economic migration requires test providers to demonstrate concurrent validity. Alignments can change over time. This can be the result of several factors, including changes in test familiarity, testing populations, test or item formats, the application of scoring rubrics, or standard setting procedures. Information regarding these issues is not publicly available for both tests, however the primary responsibility of testing organisations is to demonstrate score concordance using robust methodologies. This linking study has aggregated

data collected over several years and most recently under tightly controlled conditions to update the PTE Academic and IELTS Academic concordance table to reflect current experience of test takers and to be used by accepting institutions.

References

- Albano, A. D. (2016). equate: An R Package for Observed-Score Linking and Equating. *Journal of Statistical Software*, 74(8). <https://doi.org/10.18637/jss.v074.i08>
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching assessment (CEFR). A manual*. Strasbourg: Language Policy Division.
- De Jong, & Benigno. (2017). *Alignment of the Global Scale of English to other scales: The concordance between PTE Academic, IELTS, and TOEFL*. Pearson. Retrieved from Pearson website: <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/TeacherResources/GSE/GSE-Alignment-other-scales.pdf>
- De Jong, J., Mayor, M., & Hayes, C. (2016). *Developing Global Scale of English Learning Objectives aligned to the Common European Framework*. Pearson. Retrieved from Pearson website: <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/TeacherResources/GSE/GSE-WhitePaper-Developing-LOs.pdf>
- De Jong, & Zheng. (2016). Linking to the CEFR: Validation Using a Priori and a Posteriori Evidence. In J. Banerjee & D. Tsagari (Eds.), *Contemporary second language assessment* (pp. 83–100). London : Bloomsbury Academic.
- 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. <https://doi.org/10.1002/j.2333-8504.2010.tb02236.x>
- Eckes, T. (2012). Examinee-centered standard setting for large-scale assessments: The prototype group method. *Psychological Test and Assessment Modeling*, 54(3), 257.
- IELTS Demographic Data. (2019). Retrieved October 30, 2020, from <https://www.ielts.org/research/demographic-data>
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York: Springer-Verlag.
- Lim, G. S., Geranpayeh, A., Khalifa, H., & Buckendahl, C. W. (2013). Standard Setting to an International Reference Framework: Implications for Theory and Practice. *International Journal of Testing*, 13(1), 32–49. <https://doi.org/10.1080/15305058.2012.678526>
- Lopes, S. (2010). Test security: Defeating the cheats. *Biometric Technology Today*, 2010(4), 9–11. [https://doi.org/10.1016/S0969-4765\(10\)70085-0](https://doi.org/10.1016/S0969-4765(10)70085-0)
- North, B. (2000). *The Development of a Common Framework Scale of Language Proficiency*. Berlin, New York: Peter Lang.
- Pearson. (2010). *Aligning PTE Academic Test Scores to the Common European Framework of Reference for Languages*. Retrieved from https://pearsonpte.com/wp-content/uploads/2014/07/Aligning_PTEA_Scores_CEF.pdf
- Pearson. (2014). *FactSheet: Accurate*. Retrieved from <https://pearsonpte.com/wp-content/uploads/2014/07/AccurateFactsheet.pdf>

- 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.
<https://doi.org/10.1177/0146621604265033>
- Taylor, L. (2004). IELTS, Cambridge ESOL exams and the Common European Framework of Reference (CEFR). *Cambridge English Research Notes*, (18), 2–3.
- Zheng, & De Jong. (2011). *Establishing Construct and Concurrent Validity of Pearson Test of English Academic* [Research Note]. Pearson. Retrieved from Pearson website:
http://pearsonpte.com/wp-content/uploads/2014/07/RN_EstablishingConstructAndConcurrentValidityOfPTEAcademic_2011.pdf

Appendix 1 – Comparison of PTE Academic and IELTS Academic tests

	PTE Academic	IELTS Academic
Purpose	Both tests claim their purpose is to assess the English language proficiency required for international work, study, or immigration	
Test score use	Both PTE Academic and IELTS test scores are used in a variety of international high stakes selection decision contexts, including education, immigration, and employment decisions.	
Assessed skills	Listening, reading, speaking, writing	Listening, reading, speaking, writing
Administration	Computer-based for all four skills, including speaking.	Paper-based and computer-based offerings for listening, reading, and writing. Live interview speaking.
Test Design	Approximately 3 hours 20 item types assessing integrated skills 70 items 3 sections (Speaking & Writing, Listening, Reading)	Approximately 2 hours 45 minutes 21 item types assessing individual skills 85 items 4 sections (Listening, Reading, Writing, Speaking)
Listening	45 – 57 minutes 17 items, including: - Summarize spoken text - Multiple choice, choose multiple answer - Fill in the blanks - Highlight correct summary - Multiple choice, choose single answer - Select missing word - Highlight incorrect words - Write from dictation	30 minutes 40 items, including: - Multiple choice - Matching - Plan/map/diagram labelling - Form/note/table/flow-chart/summary completion - Sentence completion
Reading	32 – 40 minutes 15 items, including: - Reading & writing: Fill in the blanks - Multiple choice, choose multiple answers - Re-order paragraphs - Reading: Fill in the blanks - Multiple choice, choose single answer	60 minutes 40 items, including: - Fill gaps in a passage of written text or in a table - Match headings to written text to diagrams or charts - Complete sentences - Give short answers to open questions - Answer multiple choice questions

Speaking	77 – 93 Minutes 38 items, including: • Read aloud • Repeat sentence • Describe image • Re-tell lecture	11-14 minutes 3 parts: • Introduction and interview (4-5 min) • Long turn (4-3 min) • Discussion (4-5 min)
Writing	• Answer short question • Summarize written text • Essay (200-300 words)	60 minutes 2 tasks: • Describe, summarise, or explain a graph, table, chart (150+ words) • Essay (250+ words)
Scoring	All sections automatically scored by AI-based scoring engine trained by expert human judges.	Productive responses scored by expert human judges.
Productive scoring criteria	• Content • Oral Fluency • Pronunciation • Form • Development, structure and coherence • Grammar • General linguistic range • Vocabulary	• Task achievement/task response • Coherence and cohesion • Lexical resource • Grammatical range and accuracy • Fluency and coherence • Lexical resource • Grammatical range and accuracy • Pronunciation
Score Scale	Global Scale of English (GSE) 10-90	Bands 1 to 9 in half-band increments
Standard Error of Measurement (SEM)	SEM = 2.3 GSE points	SEM=0.23 of an IELTS band