

# Investigating the Relationship between Pearson PTE Scores and IELTS Bands

January 2021

Mark Elliot, Senior Research and Validation Manager  
Andy Blackhurst, Principal Validation Manager



## Summary

---

Making comparisons between scores on different tests is challenging because test products differ in their design, purpose, and format (Taylor, 2004, Lim et al, 2013) and the greater the difference in design the more problematic the exercise is. Nonetheless, test score users are often interested to know how for practical purposes results on two differing tests may compare. This report summarises the findings of a small-scale empirical study intended to explore the relationship between scores on the IELTS test and scores on the Pearson Test of English Academic (PTE Academic), conducted as part of a programme of work exploring the comparability of different approaches to test design.

Correlations between scores on the individual skills were moderately good, with the notable exception of speaking. This indicates that the differences between the IELTS face-to-face Speaking test and the automated, indirect, PTE Speaking test do indeed amount to significantly divergent test constructs. Given the differences in how the Speaking and indeed the Writing sections function, recognising institutions need to consider those aspects of test coverage as well as a simple comparison of test outcomes. The Pearson test may not always be a functional equivalent for IELTS in contexts where IELTS has historically been used.

## Context

---

Since its launch in 2009, the Pearson Test of English Academic (PTE Academic) has slowly built a profile as an English proficiency test in both academic and migration contexts. Unlike longer-established proficiency tests, it is a fully computerised test, both delivered and scored by a computer for all four skills (speaking, writing, reading, and listening). This, of course, offers significant benefits in terms of test practicality. The question that arises, however, is whether this full automation might limit and, to some degree, disrupt what the test can/should assess. The Speaking section of PTE Academic consists of relatively simple and constrained tasks such as reading aloud and repeating sentences, tasks that are intended to produce candidate speech that can be readily measured by a computer. This type of mechanical, “inauthentic” task can provide a basic level of information about a candidate’s speaking skills in a decontextualised setting, but not beyond (Chun, 2006, 2008; Galaczi, 2010; Xu, 2015). With this narrow scope, it is not sufficient to extrapolate the test results to the target domain and estimate an ability to cope with spontaneous interactions with interlocutors, a crucial component of spoken language use in real life (Galaczi and Taylor, 2018).

In addition, in-depth analyses conducted in Lu (2012) suggest that some of the traits assessed by the PTE Academic Writing section, i.e., *essay length and spelling accuracy*, are both conceptually and empirically inadequate for the assessment of academic writing ability. It is reasonable to conclude that these were employed precisely because they lend themselves to machine rating, rather than because they have a strong relationship to academic or

professional writing genres, which limits their usefulness in extrapolating from test performance to real-world language use.

The accuracy of current auto-markers for high-stakes contexts is also in question, as shown in some data-driven research by Lu (2012) and McGee (2006). Auto-markers can assess content as well as language, but the underlying mechanism in a machine's "comprehension" of meaning does not work in the same way as a human examiner's processing of a given text. In fact, the machine recognises clusters of contextual words only, not considering their logical sequence and relationship, so technically speaking, *the development of ideas and coherence of the content cannot be effectively evaluated* in the current auto-marking system (Lu, 2012; McGee, 2006).

Nonetheless, Pearson have undertaken studies comparing the performance of test takers taking both Pearson and IELTS tests in 2009 and in 2020 (with somewhat different outcomes). This paper summarises the results of two comparative studies undertaken by or for the IELTS partnership. The first of these studies (The Qualitative Study) consisted of a comprehensive qualitative comparison project undertaken by Professor Guoxing Yu (University of Bristol, UK), while the second (The Quantitative Study) was undertaken by experts from Cambridge Assessment English based on test taker performances in 2019.

In contexts where test result users are satisfied that both tests provide suitable test coverage, this report may provide useful empirical evidence to inform decisions about standard-setting.

## The Qualitative Study

---

In this study, commissioned by the IELTS partnership, comparing IELTS and PTE Academic, Yu (2020) used Kolen & Brennan's (2014) *Degrees of Similarity* (DES) framework to offer a broad comparison between the two tests. He also applied Weir's (2005) socio-cognitive framework as the basis of a detailed exploration of test task performance parameters. In addition, Yu interviewed individuals who had taken both tests in order to gain additional insight into their experiences and observations.

Yu (2020: 5) defines the four test features that form part of the DES framework as:

*Populations* refers to "to what extent are the two tests designed to be used with the same populations?".

*Constructs* refers to "to what extent do the two tests measure the same constructs?".

*Measurement characteristics/conditions* refers to "to what extent do the two tests share common measurement characteristics or conditions including, for example, test length, test format, administration conditions, etc.?" (p.498).

*Inferences* refers to "to what extent are scores for the two tests used to draw similar types of inferences?".

Note that in his study, Yu combines the Constructs and Measurement characteristics/conditions features “[G]iven the interconnectedness and overlapping” nature of these two elements.

## **Overview of the Qualitative Study**

In this section, we present some of the main findings from Yu (2020). It should be noted that the observations below offer only an overview of the extensive report presented by Yu, though the following section has been reviewed by him and acknowledged as an accurate summary of his study (Guoxing Yu, personal communication January 2021).

In the following section we review the findings by sub-skill, as was the case in the original study. In each case we begin with a number of quotations from the original report and end with our interpretation of what these might signify.

### Populations

“the two tests target the same populations of test-takers and test-score users” (p.5)

This quotation suggests that it is appropriate to compare the two tests and that we should expect that there will be some significant overlap across the tests in terms of the construct and measurement characteristics and conditions.

### Constructs and Measurement characteristics/conditions

#### *Speaking*

“IELTS and PTE Speaking tasks vary to a large extent, in terms of features of task input and conditions, expected responses, and assessment criteria and methods.” (p.28)

“We do not know the weighting of each [PTE] task, nor the percentage weighting of each skill assessed in the integrated Speaking tasks.” (p.8)

“PTE Speaking is slightly less linguistically demanding in all tasks but “retell lecture” than IELTS Speaking.” (p.27)

This indicates that the Speaking tests are very different in terms of how they assess the skill and what aspects of the skill are tested. The lack of transparency of how the PTE estimates overall ability in speaking makes comparison difficult. Issues with the well-documented failure of native speakers (BBC, 2018) on the [PTE] Speaking test point to additional issues of potential bias or inaccuracies in the auto-scoring system used. The low correlations for speaking found in this study and in the Pearson report (and in the Quantitative Study below) appear to support this implication.

#### *Writing*

“As there is no publicly available information on the weightings of each skill in the five integrated writing tasks in PTE, it is a big challenge to evaluate the cognitive and linguistic demands of the PTE Writing tasks.” (p.9)

“three [of six] PTE writing tasks (summarize written text, summarize spoken text, and write essay) ... are direct assessment of writing appropriate for academic purposes. However, to what extent the other three writing tasks (read to fill in blanks, listen to fill in blanks, with single

words; and write from dictation of a short sentence) can assess academic writing ability is questionable.” (p.9)

“PTE and IELTS may be assessing very different kinds of writing skills.” (p.33)

“Overall, PTE Writing test was considered moderately more demanding, both linguistically and cognitively, as it requires candidates to complete a range of independent and integrated tasks of different kinds of inputs and response formats under tight time control for each question ... However, the interview data as well as the high PTE scores that Interviewees 2 and 3 achieved seemed to contradict the finding from the textual analysis of the writing tasks that PTE Writing is relatively more challenging than IELTS.” (p.9)

“[Female interviewee] believed that PTE Writing was very “coachable”, as did the teacher of her PTE preparation course.” (p.33)

While it is very difficult to compare the two tests with regard to writing, there is a clear discrepancy across the PTE tasks, half of which do not appear to assess appropriate language and are clearly quite significantly below the level of difficulty expected. It also appears to be the case that the PTE Writing test can possibly be ‘gamed’ by test takers and those preparing them for the test.

The analysis of score data presented below seems to support this interpretation.

However, it should again be noted that the lack of transparency of how the PTE estimates overall ability in writing, just as with speaking, makes comparison difficult.

### *Listening*

“PTE is slightly more linguistically and cognitively demanding. This finding is different from Taylor and Chan (2015)” p.40

“easier tasks in PTE Listening which assess mainly understandings of words at a local level may well cancel out or neutralize the difficulties from the more challenging tasks such as “summarize spoken text” and “retell lecture” that require and assess global understanding of the audio inputs. As a result, it is probable that the overall linguistic and cognitive demands of IELTS and PTE may not be massively different.” p.40

“Wei and Zheng (2017) identified that “repeat sentence”, “highlight incorrect words”, “select correct summary”, and “listen-to-fill-in-blanks” are the four easiest tasks; and “write from dictation”, “repeat sentence” and “select missing words” are the three best predictors of overall listening performance. The finding on the four easiest tasks, as I predicted, was not a surprise at all because these tasks involve minimal or no language production or transformation from the source materials. However, I was rather taken by surprise that the most challenging task, “summarize spoken text” was not a good predictor at all of candidates’ overall performance in Listening. Unlike the authors who interpreted the findings by claiming that there is no association between item difficulty level and item type, I would look at this from a different, perhaps more sensible, and neutral/unbiased perspective.”

- What are the weightings of “Listening” in these independent and integrated Listening tasks?
- Could it be because the easier (more difficult) tasks make a larger (smaller) contribution to the overall Listening score?
- Does “Listening” have only a much smaller weighting than “Writing” and “Speaking” in the two more challenging summarization tasks (retell lecture, summarize spoken text)?

“Without knowing the answers to these questions with certainty, it is hard to tell the difficulty levels of the tasks and their contribution to the overall Listening score.” (p.39)

The implication here is that there is likely to be relatively little difference in the difficulty of the Listening papers across the two tests. However, they both take quite different approaches (though there are overlaps). The inconsistency of the PTE approach (with some quite difficult elements and other significantly easier elements) when coupled with the lack of transparency in reporting how the overall Listening score is arrived at makes any meaningful comparison from a measurement perspective quite difficult as it is not possible to fully interpret the findings.

### *Reading*

Yu questions the ‘fill in the blanks’ tasks in PTE, finding that “the *Reading: Fill in blanks task* is relatively easier than the texts for *Reading and Writing: Fill in blanks task*”, suggesting that it was “not very clear to me why the two slightly different tasks have different assessment focus”. (p.43)

“apart from “summarize written text” which does involve substantial reading, the other three additional tasks (read-aloud a text up to 60 words, highlight correct summary, highlight incorrect words) can only assess Reading minimally.” (p.44)

“Overall, it was found that PTE Reading is less demanding than IELTS Reading, although the reading passages are similar in many textual features.” (p.11)

The suggestion by Yu (2020: 11) that the PTE Reading paper is likely to be less demanding than the equivalent IELTS paper seems clear from his analysis. The evidence from the analysis undertaken as part of this study (see Table 2 below, p.8) appears to support this interpretation. The difference, however, is not considerable and clearly requires further study.

### Inferences

“As both tests have largely the same target populations, the inferences about test-takers’ language abilities and the decisions made based on their abilities are broadly the same: for academic study, immigration, and professional registration.” (p.6)

“However, there is no publicly available information on how the raw scores are converted to PTE scores, neither is there any information on the weightings of different items.” (p.6)

A number of test users around the world “tend to take the face value of what exam boards provide to them at the wholesale level ... many top UK universities (especially for those competitive programmes) that accept both IELTS and PTE for admission purposes ask for a higher PTE grade than what Pearson claims as equivalent to IELTS scores. There are top universities in the UK, e.g., Oxford and Cambridge, that do not accept PTE for their student admission processes. After the completion of the project, I also noticed that some universities (e.g., University of Edinburgh) were phasing out the use of PTE scores for admission purposes. The three interviewees held the view that their PTE scores are a bit inflated if their IELTS scores are held as a reference or comparison point. From my analysis of the features of the two tests in terms of their content, constructs, and measurement characteristics/conditions (the findings are summarized below), I am in agreement with the judgement of the three interviewees and the universities.” (p.6)

It would seem that while it is feasible that the inferences drawn from test performance are generally similar for both tests, there are a number of issues that test score users should take into consideration when deciding on which test is suitable for use in their specific context. These can be summarised as:

- There appears to be a difference in difficulty across the two tests. The PTE is felt by at least two stakeholder groups (test takers and universities) to be easier (i.e. less cognitively and linguistically challenging). This suggests that receiving organisations and governments should review the cut scores they currently apply to PTE.
- There is a significant lack of transparency in the way in which PTE sections and overall scores are calculated.
- There is some evidence from other sources (e.g. the failure of native speakers to achieve the score demanded of them for PTE Speaking) that there is a significant bias built into the auto-scoring system used. No specifications are available for the training data used for the system, and no estimate of the accuracy of the transcription system (from voice to text) is provided by Pearson – the industry norm is referred to as the Word Error Rate (WER).

## Conclusions and Recommendations from this Study

---

Yu (2020: 54) suggests that a broader analysis of the two tests is urgently required, suggesting that the “IELTS Partners have the responsibility to produce their version of the IELTS/PTE equivalence table, but beyond a comparison between the overall scores/bands”. The summary of his work included here when combined with the analysis of the score data presented below marks the first stage of the comprehensive programme of comparison prompted by this suggestion.

## The Quantitative Study

---

The data used in this study were obtained by Catalyst Research, an independent research firm of Perth, Australia, and Macquarie University International College English Language Centre as part of a survey of test taker experiences with different tests. Score information was obtained from 523 test takers who had taken both tests (IELTS Academic and PTE Academic) within 90 days of each other. The majority had taken IELTS in Australia and represented a range of nationalities/first language backgrounds, including Chinese, Indonesian, Filipino and Polish, while smaller numbers had taken IELTS in Hong Kong, Pakistan and the UK. One hundred and fifteen participants provided only their overall scores, so analysis at individual skill level is based on just 408 test takers. The first analysis undertaken was a simple correlation between performance on the two tests, i.e. how far do they agree in their rank-ordering of the test takers? This is of interest because it points to the extent to which the tests can be regarded as testing the same *construct* (the range of performances that the test’s design seeks to assess and the tasks employed to do this).

It is interesting to compare these correlations with the broadly similar correlations quoted in Pearson (2020) of 0.74 Overall with 0.66 for Listening, 0.68 for Reading, 0.60 for Writing and 0.42 for Speaking. The correlation is strongest in terms of the overall test result and for three components the correlation is moderately (though not very) strong, which would be consistent with their being essentially similar in what they measure. In both studies the correlation between Speaking results is much weaker, implying that the tests are diverging in terms of what they are testing. This is confirmation that the IELTS face-to-face Speaking test (which has historically provided the normative benchmark for testing speaking) and the automated, indirect, PTE Speaking test do not yield similar results and that the Pearson test **is not always a functional equivalent for IELTS in contexts where IELTS has historically been used.**

**Table 1: Correlations between scores on IELTS and scores on PTE**

Component	Pearson correlation This Study	Pearson correlation PTE Study (2020)
Overall	0.70	0.74
Listening	0.66	0.66
Reading	0.60	0.68
Speaking	0.44	0.42
Writing	0.62	0.60

Equipercentile linking with presmoothing, as described in Kolen & Brennan (2014), was chosen as the analytic method for further analysis of the data. This approach to smoothing is advantageous in that indices are available for evaluating the goodness of fit and therefore of the linking. The linking was carried out using the RAGE-RGEQUATE software (Zeng, Kolen, Hanson, Cui, & Chien, 2004). The output may be found in the appendices and is summarised in table 2 below.



**Table 2: Putative alignment of IELTS bands and Pearson PTE scores (Equipercntile Equating)**

IELTS	PTE Overall	PTE Listening	PTE Reading	PTE Speaking	PTE Writing
8.5	88.1	84.7	83.7	85.5	89.5
8	82.3	79.4	78.4	80.9	89.4
7.5	74.6	73.9	73.7	75.3	87.5
7	66.3	66.2	67.8	65.3	82.3
6.5	58.5	56.8	60.6	53.5	74.1
6	51.6	48.1	53.5	46.2	62.2
5.5	45.4	42.7	47.9	42.2	51.0
5	40.8	40.2	43.0	40.2	43.1

It should be noted how much the Writing alignment diverges from the others. This is true even at the mid-range of ability but the difference is greater at the higher levels. At the same time, it can be seen that there is less score progression at the top of the scale in Writing, so that while the increments between half bands in other skills approximates to 5 scale points on the Pearson scale, only 2 scale points on the Pearson scale separate band 7.5 and band 8.5 in Writing.

## Conclusions and Recommendations from this Study

---

Attention has been drawn to the caveats that must be entered as to the comparability of test results, given the different judgements made in the test design process as regards testing the productive skills. This holds true for the comparison of any two tests but in this case the divergence in test outcomes is readily apparent. Individual institutions should consider a range of evidence when setting standards for their specific purposes, as the range of activities sampled by different tests (and the depth in which they do so) differs and so the applicability of scores may vary, depending on the range of activities in which applicants will typically be engaged.

Nonetheless, Table 2 above provides an empirical basis for institutions wishing to compare applicants who have taken the different tests in terms of identifying the broadly defined level of ability at which they are placed. As noted above, the correlation between outcomes is stronger for the test overall and so pragmatically, the differences in how the two tests elicit and process behaviours in the realm of individual skills can be managed to some extent by focusing on the overall outcome score across the skills.

## Overall Conclusions and Recommendations

---

Despite the caveats that have been pointed out in this report, the multi-method approach taken here allows us to draw a number of conclusions from the studies. The most obvious of these are as follows:

- The PTE Writing paper would appear to be significantly more leniently scored than the equivalent IELTS paper.
- This issue impacts on the overall scores awarded for what would be similar levels of performance on the two tests compared here.
- There appears to be a significant difference in the way in which the speaking skill is tested and scored across the two tests. The tendency of the Speaking test to result in quite different profiles is indicated by the correlation coefficients presented both by IELTS and Pearson (see Table 1). The suggestion here is that test score users should carefully review the two Speaking tests to identify the most appropriate for their context.
- Without a clear understanding of how the PTE test subsection and overall scores are calculated it is recommended that test users take great care when making decisions as to their appropriateness. The lack of transparency, which is conspicuous in Yu's discussion of the PTE Listening paper but which is actually present across all of the PTE Academic reporting system, seriously impacts on test score interpretation and as such may undermine any validity claims made for the test.
- Without an understanding of the data upon which the scoring systems have been trained and considering the negative experience of English native speaker test takers reported above, it is possible that unintended bias is present in the system. This can have a significant impact on the scores awarded to test takers from unrepresented or under-represented groups. In addition, where particular accents are associated with negative scores in the training data there is a likelihood of negative bias (the opposite will be the case when high performance samples are used), (see Buolamwini & Gebru, 2018 and Gebru et al, 2020).

While the concordance table presented here (Table 2) tells us quite a lot about the relationship between the two tests, the qualitative data offers a vital additional insight. The IELTS partnership therefore recommends that test users refer to the table when making decisions, but at the same time we believe that it is necessary to look beyond the numbers to understand more fully the strengths and weaknesses of tests that are presented to them. Yu (2020: 7) tells us that

*There are top universities in the UK, e.g., Oxford and Cambridge, that do not accept PTE for their student admission processes. After the completion of the project, I also noticed that some universities (e.g., University of Edinburgh) were phasing out the use of PTE scores for admission purposes.*

This suggests that top level institutions in the UK are beginning to take this more critical approach.

## References

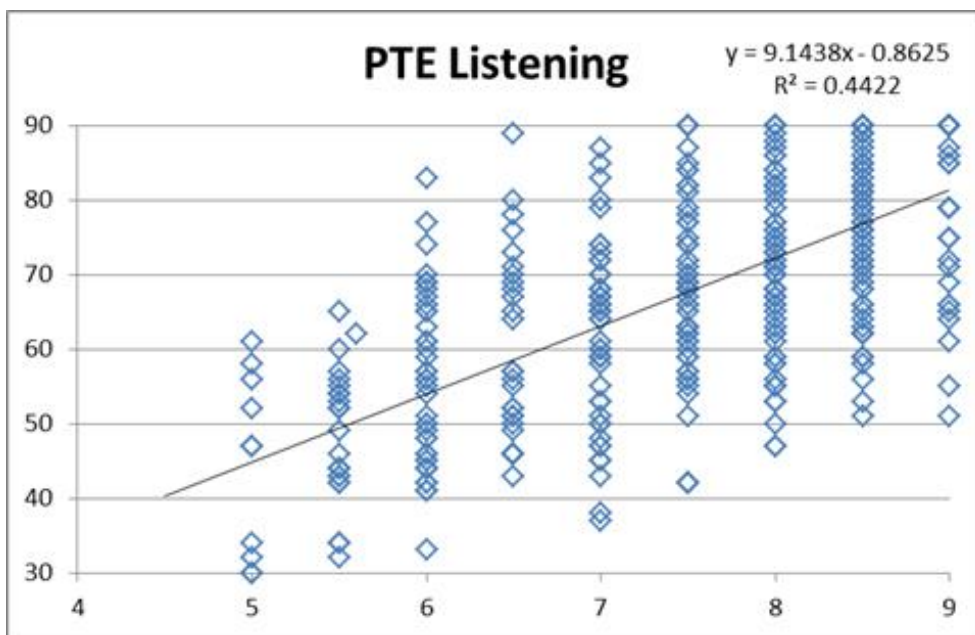
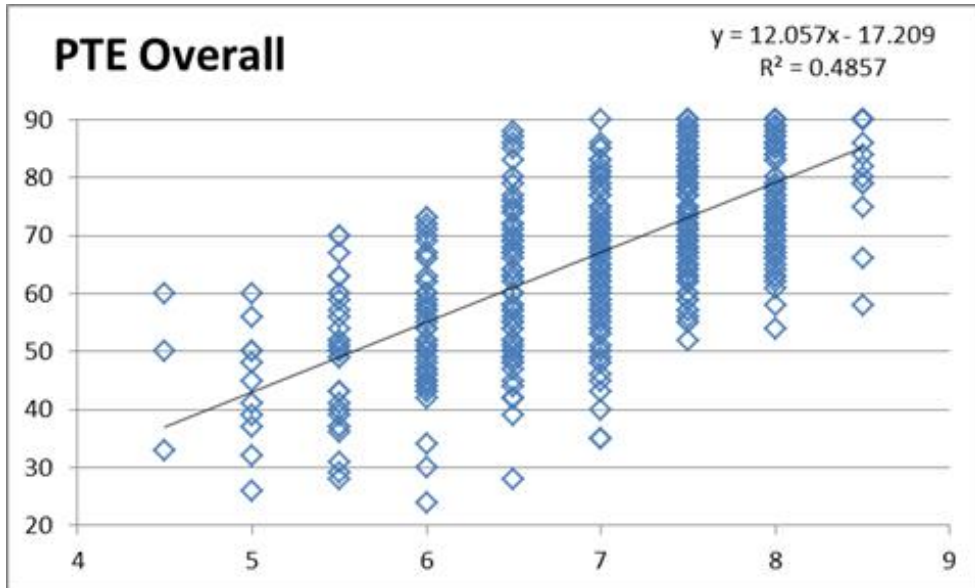
---

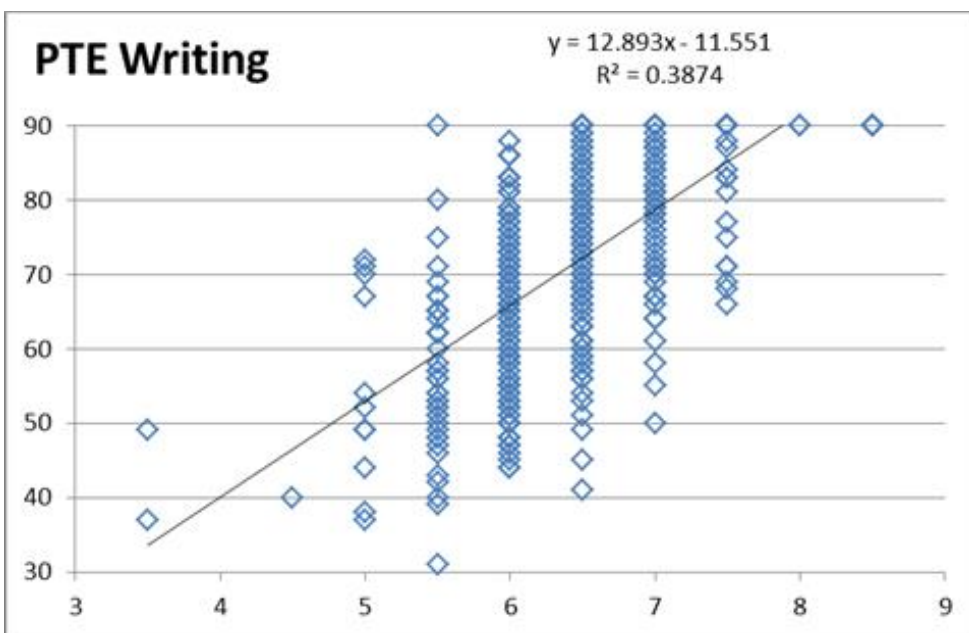
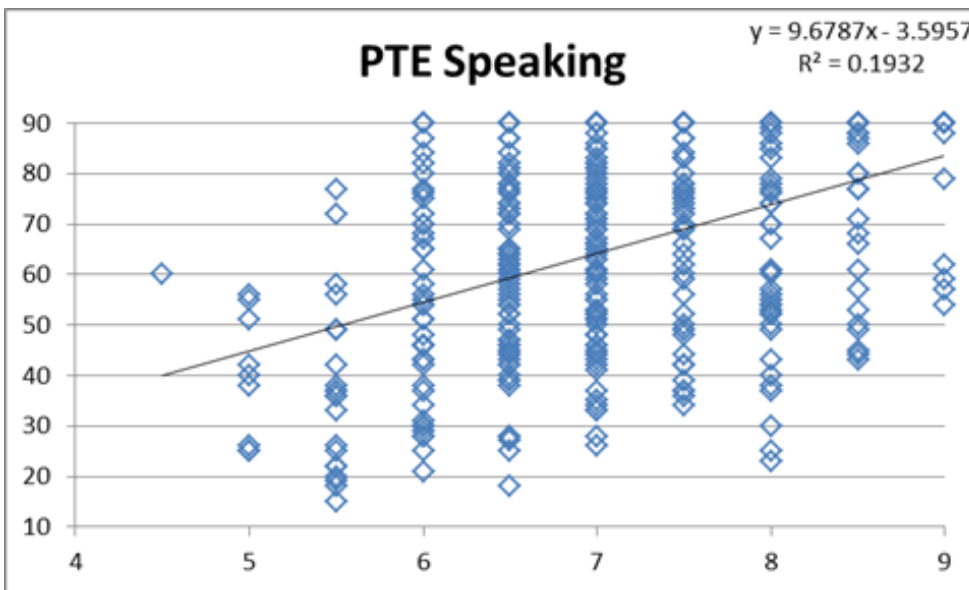
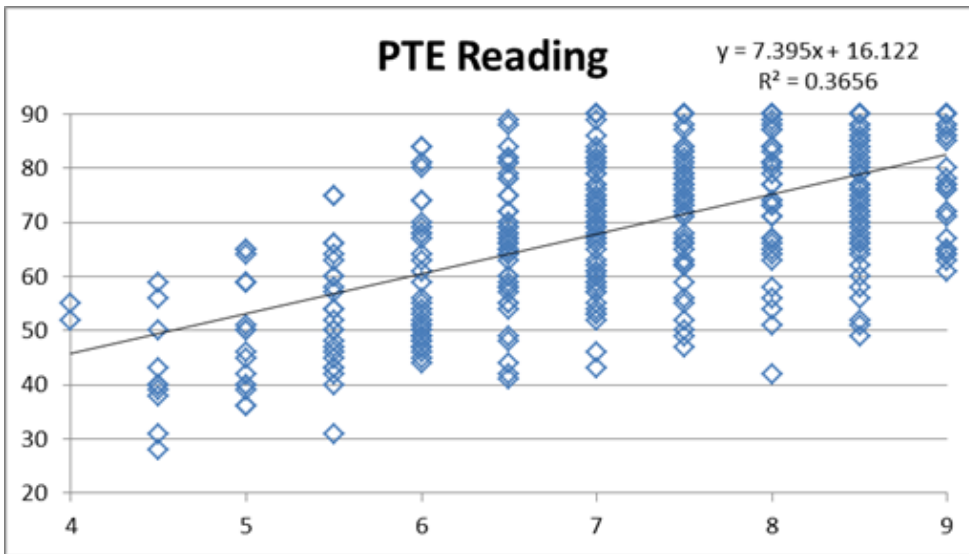
- BBC (2018). *Irish vet's English fails Australian computer visa test*. BBC News Website accessed on January 5, 2021 from: <https://www.bbc.co.uk/news/world-europe-40868315>
- Buolamwini, J. & Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Proceedings of Machine Learning Research 81: 1–15. Accessed on January 5, 2021 from: <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>
- Chun, C. W. (2006). An analysis of a language test for employment: The authenticity of the PhonePass test. *Language Assessment Quarterly*, 3(3), 295–306.
- Chun, C. W. (2008). Comments on “evaluation of the usefulness of the *Versant for English Test*: A response”: The author responds. *Language Assessment Quarterly*, 5(2), 168–172.
- Clesham, R. & Hughes, S. R. (2020). 2020 Concordance Report: PTE Academic and IELTS Academic. Accessed on January 14, 2020 from: <https://pearsonpte.com/wp-content/uploads/2020/12/2020-concordance-Report-for-research-pages.pdf>
- Elliot, M. (2020). IELTS and PTE equipercentile equating of scores using Rage-Equate software with pre-smoothing. Cambridge Assessment English Internal Report, January 2020.
- Galaczi, E. D. (2010). Face-to-face and computer-based assessment of speaking: Challenges and opportunities. In L. Araújo (Ed.), *Computer-based assessment of foreign language speaking skills* (pp. 29–51). Luxembourg: European Union.
- Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, 15(3), 219–236.
- Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H. & Crawford, K. (2020). Datasheets for datasets. Working Paper accessed January 5, 2020 from: <https://arxiv.org/pdf/1803.09010v7.pdf>
- Kolen, M. J., & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking: Methods and Practices*. Springer Science & Business Media.
- Lim, G. S, Geranpayeh, A., Khalifa, H. and Buckendahl, C. W. (2013). Standard setting to an international reference framework: Implications for theory and practice, *International Journal of Testing*.
- Lu, L. F. (2012). A validation framework for automated essay scoring systems (Unpublished doctoral dissertation). University of Wollongong, Wollongong, Australia.
- McGee, T. (2006). Taking a spin on the Intelligent Essay Assessor. In P. F. Ericsson & R. Haswell (Eds.), *Machine scoring of student essays, truth and consequences* (pp. 78–92). Logan, UT: Utah State University Press.
- Pearson plc (2020) *Concordance Report: PTE Academic and IELTS Academic*

- Taylor, L. (2004). Issues of test comparability, *Research Notes* 15, 2–5.
- Taylor, L., & Chan, S. (2015). *IELTS Equivalence Research Project (GMC 133)*.  
[https://www.gmc-uk.org/-/media/documents/GMC\\_Final\\_Report\\_\\_\\_Main\\_report\\_\\_extended\\_\\_\\_\\_Final\\_\\_\\_13May2015.pdf\\_63506590.pdf](https://www.gmc-uk.org/-/media/documents/GMC_Final_Report___Main_report__extended____Final___13May2015.pdf_63506590.pdf)
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Palgrave Macmillan.
- Xu, J. (2015). *Predicting ESL learners' oral proficiency by measuring the collocations in their spontaneous speech* (Unpublished doctoral dissertation). Iowa State University, Ames, USA.
- Yu, G. (2020). *IELTS and PTE: Degrees of Similarity*. Commissioned report for the IELTS Partners.
- Zeng, L., Kolen, M. J., Hanson, B. A., Cui, Z., & Chien, Y. (2004). RAGE-RGEQUATE [Computer software]. Iowa City: University of Iowa.

# Appendix 1

Scatterplots of scores obtained on Pearson as compared to IELTS

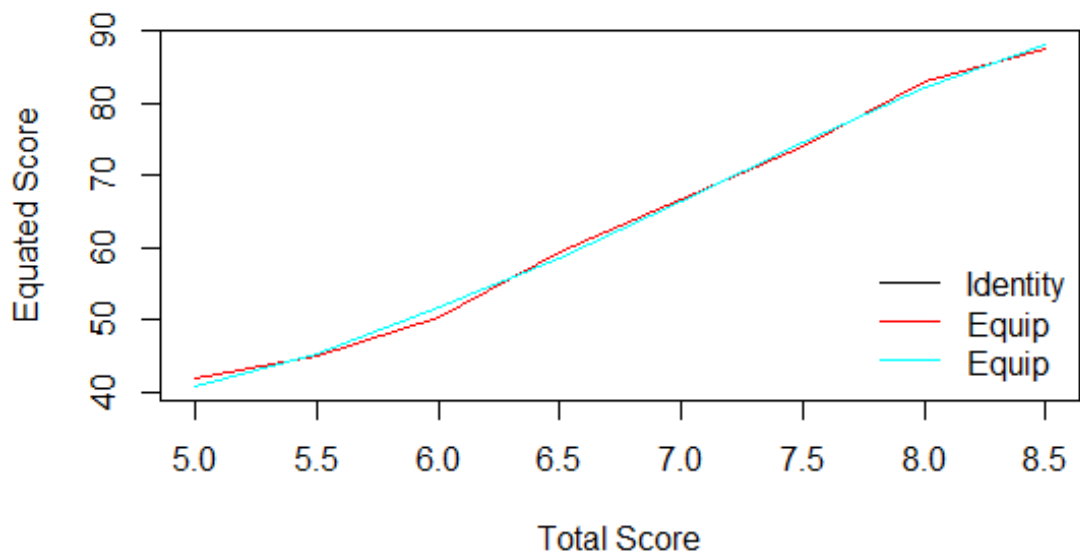




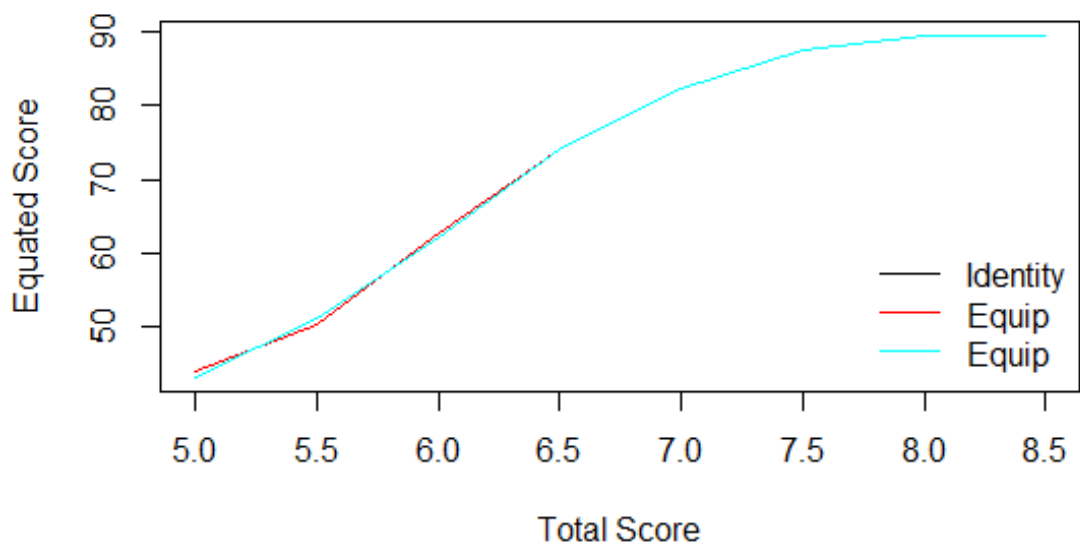
## Appendix 2 Equipercentile graphs

---

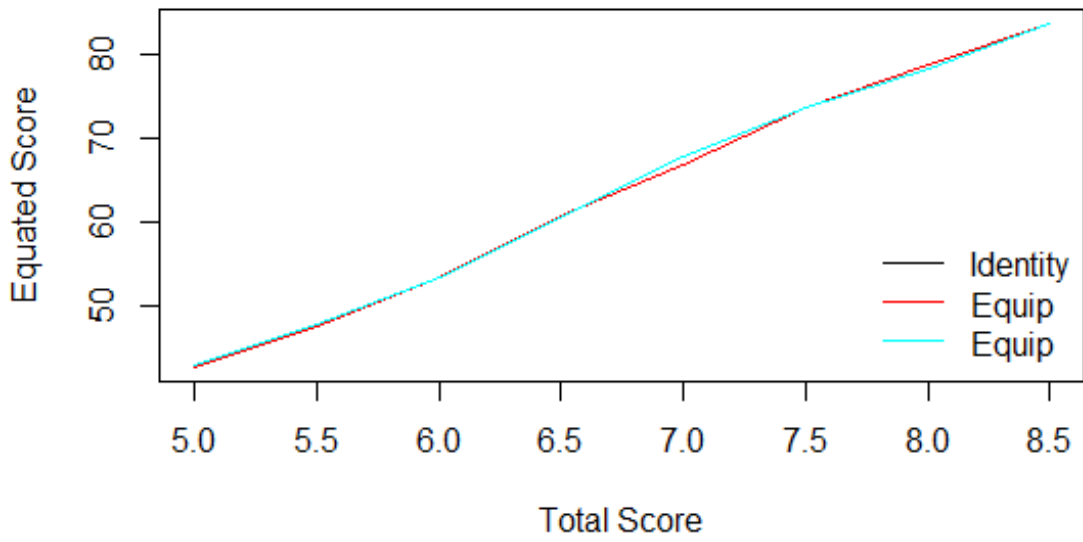
Overall



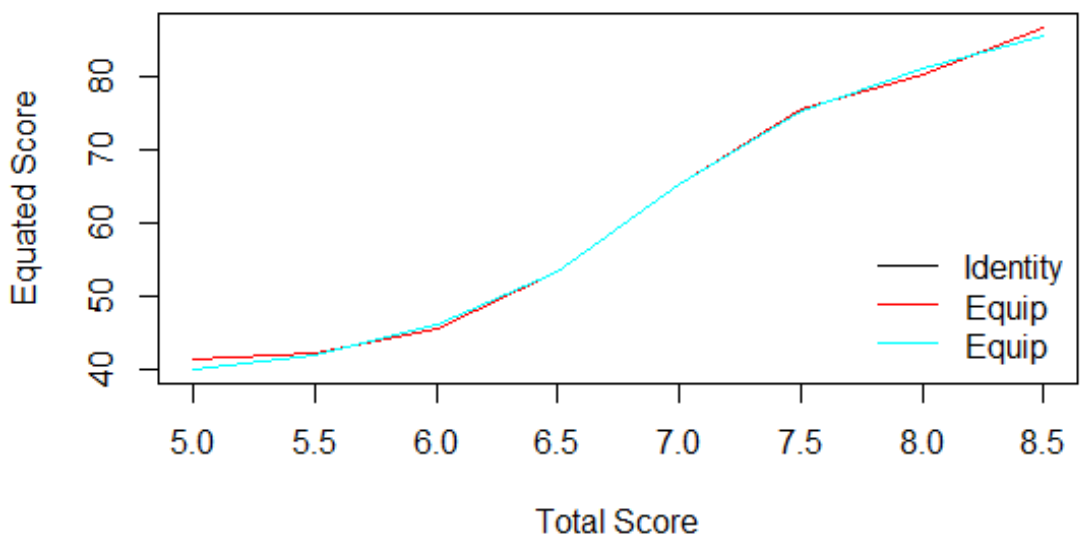
Writing



## Reading

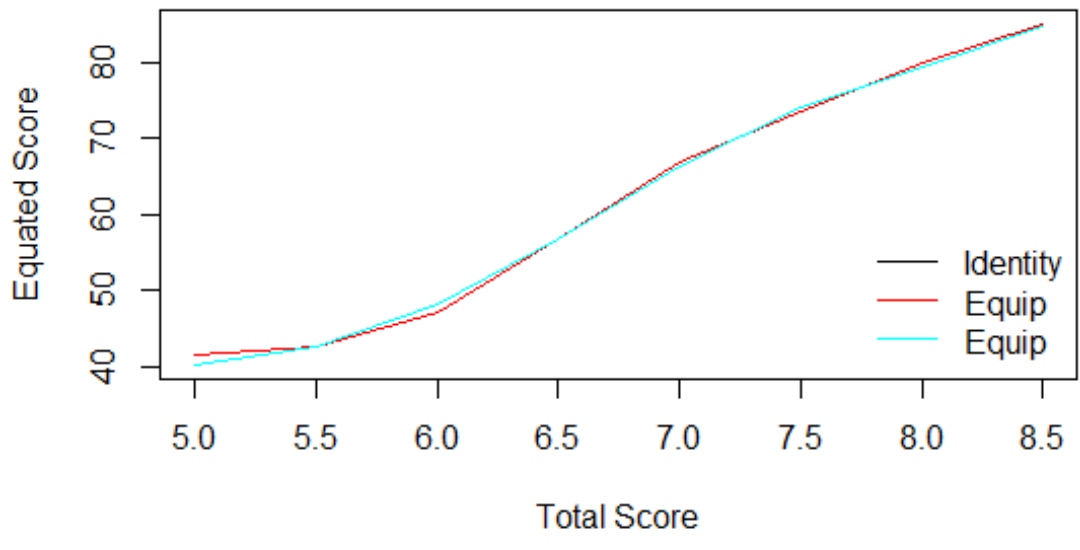


## Speaking





# Listening



## Appendix 3 Concordance tables

---

### Overall

	scale	yx	se	se.b
1	5.0	40.76146	0.8628093	0.7845668
2	5.5	45.35398	1.2665579	1.0969325
3	6.0	51.58694	1.2318055	1.1678655
4	6.5	58.53999	1.2932963	1.2280964
5	7.0	66.27297	1.2587240	1.1635235
6	7.5	74.55021	1.0962364	0.9920956
7	8.0	82.30825	0.9493683	0.8584286
8	8.5	88.11916	0.7309762	0.6512508

### Writing

	scale	yx	se	se.b
1	5.0	43.13323	1.71946867	1.39411761
2	5.5	50.97456	1.38290617	1.20679319
3	6.0	62.15329	1.49294727	1.32356210
4	6.5	74.06259	1.18313320	1.07792821
5	7.0	82.32697	0.93661065	0.87295713
6	7.5	87.50599	0.97416557	0.70536971
7	8.0	89.36798	0.22983379	0.29954861
8	8.5	89.49843	0.02360109	0.01744959

### Reading

	scale	yx	se	se.b
1	5.0	42.99891	1.410367	1.2067798
2	5.5	47.89908	1.501075	1.4593999
3	6.0	53.49646	1.655884	1.5682711
4	6.5	60.55533	1.666954	1.6147487
5	7.0	67.84451	1.491879	1.4686746
6	7.5	73.73299	1.252517	1.2428936
7	8.0	78.35382	1.096181	1.0768625
8	8.5	83.69480	1.090230	0.9061987

### Speaking

	scale	yx	se	se.b
1	5.0	40.15496	0.6589449	0.5723585
2	5.5	42.17077	1.0516333	0.9467973
3	6.0	46.20474	1.2438123	1.1304001
4	6.5	53.46676	1.6919827	1.6524944
5	7.0	65.25109	2.2849466	2.1773222
6	7.5	75.32197	1.6737388	1.6209173
7	8.0	80.90768	1.3292715	1.2748341
8	8.5	85.50931	1.3106047	1.0289886

### Listening

	scale	yx	se	se.b
1	5.0	40.15496	0.6589449	0.5723585
2	5.5	42.17077	1.0516333	0.9467973
3	6.0	46.20474	1.2438123	1.1304001
4	6.5	53.46676	1.6919827	1.6524944
5	7.0	65.25109	2.2849466	2.1773222
6	7.5	75.32197	1.6737388	1.6209173
7	8.0	80.90768	1.3292715	1.2748341
8	8.5	85.50931	1.3106047	1.0289886