

December 2017

# A Standard Setting Study to Establish Concordance between the Pearson Test of English Academic (PTE A) and the Canadian Language Benchmarks (CLB)

## Summary

---

Written by  
Glyn Jones, Dr. John HAL de Jong,  
Dr. Ying Zheng and David Booth

Reviewed by  
Andrea Strachan

---



Pearson



## Aim

The aim of the study was to establish a table demonstrating the degree of concordance between PTE Academic (PTE A) scores and the Canadian Language Benchmarks (CLB) (CIC, 2012) four to ten for Listening, Reading, Speaking and Writing.

## Panel

A panel was recruited by advertising on appropriate regional, national and international discussion lists and through contacts Pearson Canada has with appropriate institutions. Panel members were experienced ESL practitioners with a thorough theoretical understanding of the CLB and practical experience in their application to English language instruction, curriculum development or assessment. Panellists were asked to supply key demographic and professional data as it was important that they represented Canada's geographic diversity as well as the range of educational establishments where the CLB are used. The key criterion (critical for the success of the project) was, however, their familiarity with the CLB.

<b>Panel Characteristic</b>	<b>Category</b>	<b>%</b>	<b>(N)*</b>
Gender	Female	91	21
	Male	9	2
Job Role	Instructor (Teacher, Trainer)	61	14
	Head of Department/Specialist	13	3
	Director	13	3
	(Assistant) Professor	13	3
Years of Experience with CLB	Under 5	9	2
	5 to 9	43	10
	10 to 15	26	6
	More than 15	13	3
	Not known	9	2
Canadian Provinces	Ontario	39	9
	Quebec	4	1
	British Columbia	9	2
	Alberta	17	4
	Nova Scotia	9	2
	Newfoundland and Labrador	-	-
	Saskatchewan	13	3
	Manitoba	9	2
	New Brunswick	-	-
	Prince Edward Island	-	-
Affiliation	University	35	8
	College	30	7
	Language Instructions for Newcomers to Canada (LINC)	26	6
	Other Institution	9	2

## Methodology

The workshop took place in Toronto in September 2017 with panellists (each of whom had been given a randomly assigned secret number) entering their judgments individually into booklets. Prior to the workshop, panellists received an overview of PTE A in a PowerPoint presentation, as well as links to online sources of information about the test. For Listening and Reading, a set of test items was prepared representing a cross section of the item types used to assess each skill in PTE A while for Speaking and Writing, a set of test taker responses obtained in the course of field testing or from live test administration was provided.

During the first day of the workshop, the standard setting facilitator provided an introduction to the test, focusing on the particular item types to be used in the workshop. Examples of these item types appeared as they would to a test taker, with audio playback where applicable. For Reading and Listening, panellists were asked to consider, "What CLB best describes a learner who has a 50% chance of answering this item correctly?" and for Speaking and Writing, "What CLB best describes the English language proficiency of the speaker?" The scoring criteria for each of these item types were also explained and the panellists reminded that although PTE A awards separate scores for each of the four skills, certain item types (integrated items) test more than one skill.

A separate standard setting workshop was conducted for each of the four skills, with the sequence of steps substantially the same for each workshop. For each skill, a booklet was prepared containing the set of test items (and, in the case of Writing, test taker responses) to be used in the workshop. The booklet was used to present items (and responses) to the panellists and to record their judgments, in each case in the form of a number between 1 and 12 corresponding to a CLB. Items or responses were presented in random order irrespective of empirical difficulty or item type so that the panellists would consider each item or response independently rather than by comparison with similar items or responses.

Two rounds of judgments were collected for each skill. In the first round, panellists submitted their judgments individually in the appropriate pages of their booklets, which were then collected and the judgments transferred to a spreadsheet for analysis in the second round. In the second round, panellists were shown the first round results from the whole panel enabling each to see how their own judgments compared with those of the group as a whole. Items displaying the greatest variance were selected for discussion, and while they were not obliged to revise their first round judgments, panellists were given the opportunity to do so in the light both of the collective results and of the discussion. The panellists then submitted their second round judgments in their booklets, which were collected again and the revised judgments transferred to a spreadsheet for analysis.

## Analysis

Before analysis could begin, the data was cleaned by removing individual ratings if they differed by more than 1.5 CLB levels from the average rating given by the panel overall, by removing individual ratings if their correlation with the average of all panel members was less than .5, and by removing items with low levels of agreement among panel members. After cleaning, the degree of agreement amongst raters was estimated and two sets of regressions calculated using two different methods.

The first method involved using both sets of panel ratings along with the known difficulty values and scores of test takers.

CLB	PTE A R	PTE A W	PTE A L	PTE A S	Overall
12	90	87	90	90	89
11	84	79	89	85	84
10	76	71	79	76	76
9	67	64	70	68	67
8	59	56	61	59	59
7	51	48	51	50	50
6	43	40	42	42	42
5	35	33	32	33	33
4	26	25	23	24	25

We can see a reasonable progression of difficulty through the levels with similar scores for the different skills. Levels 11 and 12 are included in the analysis as panellists were allowed to choose any level of the CLB during the standard setting activity. Inclusion of these levels again shows a progression to very high levels of language proficiency.

The second method used the data available from the concordance of PTE A with IELTS (Zheng and De Jong 2011). This data represented a sample of nearly 2500 self-reported scores across IELTS versions. This data was used to triangulate the relationship between the CLB and PTE A, IELTS and PTE A, and IELTS and the CLB. Each of the datasets have been independently collected and analysed. The CLB and PTE A relationship is as described in this document based on a standard setting exercise; the relationship between IELTS and PTE A is based on test takers data collected as part of the PTE A Concordance Validity study referenced above; the IELTS and CLB data is taken from the published score equivalence tables. Using these three independent sets of data gives a high degree of confidence in the relationship between the three measures as detailed below.

CLB	IELTS Reading	IELTS Writing	IELTS Listening	IELTS Speaking	IELTS Overall	PTE A Overall
12	9.0	9.0	9.0	9.0	9.0	86
11	8.5	8.0	9.0	8.0	8.4	82
10	8.0	7.5	8.5	7.5	7.9	77
9	7.0	7.0	8.0	7.0	7.3	69
8	6.5	6.5	7.5	6.5	6.8	62
7	6.0	6.0	6.0	6.0	6.0	50
6	5.0	5.5	5.5	5.5	5.4	41
5	4.0	5.0	5.0	5.0	4.8	33
4	3.5	4.0	4.5	4.0	4.0	26

We then compared the outcomes from the two methods. Table 4 indicates that the outcomes from the standard setting study are very close to the scores as predicted by the IELTS/PTE A concordance. This provides strong evidence that the scores set in the workshop are a true reflection of the score required on PTE A at the different levels of the CLB.

CLB	Method 1	Method 2	Difference
	PTE Academic (GSE)	PTE Academic (GSE)	Method 1 - Method 2
12	89	86	-3
11	84	82	-2
10	76	77	2
9	67	69	2
8	59	62	3
7	50	50	0
6	42	41	-1
5	33	33	0
4	25	26	2

## Conclusion

The panellists (all of whom responded) were invited to give feedback on their experience of the standard setting workshop by responding to an online survey. Interestingly, and for all four skills, the highest levels of agreement were reflected in feedback regarding confidence, suggesting that even those who experienced difficulty relating some of the items to the CLB were nevertheless able to arrive at decisions that they felt comfortable with. In general, the panellists' feedback indicated that while they found the experience challenging, they also found it interesting and rewarding.

The standard setting workshop was conducted following established methodologies as described in the methodology sections above. The results were analysed using standard statistical methods and show a consistent relationship with other measures.

Accordingly, the outcomes of the workshop provide a robust basis for establishing the cut-off points on the PTE A scale for different levels on the Canadian Language Benchmarks.

## References

- CIC, Canada. (2012). *Canadian Language Benchmarks: English as a Second Language for Adults*. Ottawa: Centre for Canadian Language Benchmarks.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting : a guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: SAGE.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 433–470). Westport CT: American Council on Education Praeger.
- Zheng, Y. & De Jong, J.H.A.L. (2011). Establishing the construct and concurrent validity of Pearson Test of English Academic. *Pearson Research summaries and notes*: [www.pearsonpte.com/organizations/researchers/research-notes](http://www.pearsonpte.com/organizations/researchers/research-notes)



