CELPIP Annual Report of 2019 Test Takers

Paragon Testing Enterprises
110-2925 Virtual Way, Vancouver, BC V5M 4X5
research@paragontesting.ca

CELPIP®

1 Description of the Test

1.1 General Description

The Canadian English Language Proficiency Index Program® (CELPIP Test) is a standardized test of general English language proficiency. It is administered by Paragon Testing Enterprises Inc., a subsidiary of The University of British Columbia.

The purpose of the CELPIP Test is to provide a valid and reliable measurement of a test taker's English abilities in a variety of everyday situations, such as communicating with co-workers and superiors in the workplace, interacting with friends, understanding newscasts, and interpreting and responding to written materials. The CELPIP Test is designated by Immigration, Refugees and Citizenship Canada (IRCC) for permanent resident status in Canada and Canadian citizenship. The CELPIP Test is also accepted by a number of post-secondary institutions and professional associations as proof of English language proficiency for academic and professional purposes.

Paragon is committed to upholding the highest standards in educational measurement. All parts of the CELPIP Test are written following specified guidelines, and results are closely monitored to ensure that they are accurate and informative. Paragon works closely with test centres to make certain that the CELPIP Test is administered in a way that is fair and accessible but also secure, ensuring that the CELPIP Test is available to all individuals who wish to take the exam and that the scores issued are defensible.

1.2 Test Format

There are two versions of the CELPIP Test: the CELPIP-General Test and the CELPIP-General LS Test. Individuals who take the CELPIP-General Test are assessed on all four components. Individuals who take the CELPIP-General LS Test are assessed on the Listening and Speaking components.

Table 1 describes the format and content of each test component.

Table 1: Format and Content of the CELPIP Test

Component	Time Allotted	Item Description	Number of Items
Listening	53 - 60 minutes	Test takers listen to seven passages and answer comprehension questions. The listening passages cover topics in daily conversation, problem-solving, news items, discussions, and viewpoints.	Minimum of 38
Reading	47 - 55 minutes	Test takers read several passages and answer comprehension questions. The reading passages engage the test takers in understanding correspondence, interpreting a diagram, and reading for viewpoints.	Minimum of 38
Writing	55 - 60 minutes	Test takers write an email and write a response to survey questions.	2 Tasks
Speaking	15 - 20 minutes	Test takers speak to give advice, talk about personal experiences, describe scenes, make predictions, compare and persuade, deal with difficult situations, express opinions, and to describe an unusual situation.	8 Tasks

2 Scoring and Reporting of Results

The CELPIP Test has been designed to assess the English language ability of test takers in general social, educational, and workplace contexts.

The multiple-choice items in Listening and Reading Components are computer-scored. Each correct answer contributes proportionately to the final score, and there are no points deducted for wrong answers. The Writing and Speaking components are each evaluated by at least three Paragon-certified raters according to a scale established by Paragon.

Test takers receive a score report that provides a score for each component. The CELPIP Test component scores are reported on 11 bands: M, 3 - 12.

The CELPIP test scores have been calibrated against the Canadian Language Benchmark (CLB) levels. Table 2 shows each CELPIP level and its corresponding description, with the CLB level equivalencies included in the third column.

Table 2: Interpretation of CELPIP Test Scores

CELPIP Level	CELPIP Descriptor	CLB Level
NA	Not Administered: test taker did not receive this test component	/
M	Minimal proficiency or insufficient information to assess	1, 2
3	Some proficiency in limited contexts	3
4	Adequate proficiency for daily life activities	4
5	Acquiring proficiency in workplace and community contexts	5
6	Developing proficiency in workplace and community contexts	6
7	Adequate proficiency in workplace and community contexts	7
8	Good proficiency in workplace and community contexts	8
9	Effective proficiency in workplace and community contexts	9
10	Highly effective proficiency in workplace and community contexts	10
11	Advanced proficiency in workplace and community contexts	11
12	Advanced proficiency in workplace and community contexts	12

When interpreting a CELPIP Test score report, it is important to remember that the CELPIP Test estimates test takers' true proficiency by approximating the kinds of tasks that they may encounter in the course of their regular daily lives, study, or at work. There are, inevitably, small mismatches between the tasks that test takers complete as part of the test and the tasks that they have to perform in a specific context. Also, temporary factors unrelated to test takers' true proficiency, such as fatigue, anxiety, or illness, may affect their CELPIP test results.

3 Test-Taking Population

This section presents an overview of the test takers who took the CELPIP Test in 2019.

Figure 1 shows the distribution of test takers for the CELPIP-General and the CELPIP-General LS Tests in the total test-taker population.

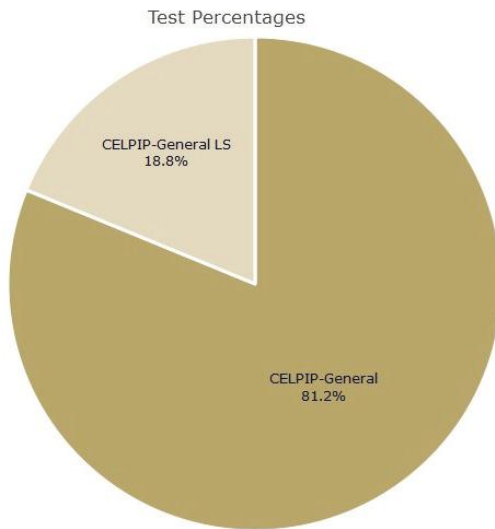**Figure 1: Distribution of CELPIP Test Takers by Test Type**

Figure 2 and Table 3 present the distributions of test takers' self-reported purpose for taking the test and age, respectively. As shown in Figure 2, the most commonly reported purpose for taking the CELPIP Test is for immigration (IRCC = 92.2%).

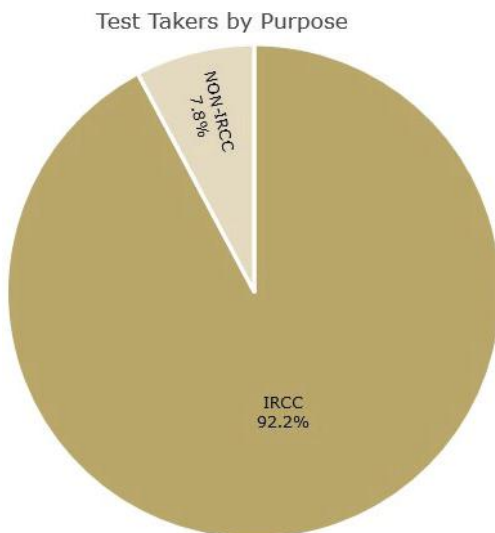**Figure 2: Distribution of CELPIP Test Takers by Purpose**

Table 3 shows that, among the test takers who registered for the CELPIP Test, the two age groups—26 to 30 (27.5%) and 30 to 35 (24.4%)—were the largest. Compared with those who took the CELPIP-General Test, more CELPIP-General LS test takers reported to be older than 40 (16.2% for CELPIP-General Test vs. 30.9% for CELPIP-General LS Test).

Table 3: Percentage of CELPIP Test Takers by Age

Age Group	% of Total Test Population	% of CELPIP-G Test Population	% of CELPIP-LS Test Population
< 20	0.3%	0.2%	0.6%
20-25	11.2%	12.6%	5.3%
26-30	27.5%	30.7%	13.9%
31-35	24.4%	24.4%	24.8%
36-40	17.6%	16.0%	24.6%
> 40	18.9%	16.2%	30.9%

4 Test Statistics

4.1 Score Descriptions

Tables 4 presents the mean and standard deviation for each component of the CELPIP-General and CELPIP-General LS Tests. The mean score is the simple average of all test takers' scores and the standard deviation (SD) quantifies how scores are spread out from the mean. Tables 5a and 5b show a more detailed breakdown of the percentage of test takers who received each component band score on the CELPIP-General and CELPIP-General LS Tests.

Table 4: CELPIP-General and CELPIP-General LS Test Scores

Test	Component	Mean	Standard Deviation
CELPIP-General	Listening	8.14	2.55
	Reading	8.21	2.97
	Writing	7.08	1.34
	Speaking	7.05	1.99
CELPIP-General LS	Listening	6.39	2.62
	Speaking	5.91	1.72

Note: To calculate the mean and standard deviation, we treated the band level M as 2.

Table 5a: Score Distributions on the CELPIP-General Test

Band	Listening	Reading	Writing	Speaking
M	0.7%	3.2%	3.2%	0.3%
3	2.7%	3.4%	3.4%	0.9%
4	6.5%	6.8%	6.8%	4.8%
5	7.4%	7.4%	7.4%	12.9%
6	9.7%	10.2%	10.2%	28.5%
7	11.2%	10.0%	10.0%	17.4%
8	19.8%	10.4%	10.4%	16.9%
9	8.4%	12.1%	12.1%	5.7%
10	10.2%	7.2%	7.2%	3.6%
11	12.1%	7.3%	7.3%	4.9%
12	11.3%	22.1%	22.1%	4.0%

Table 5b: Score Distributions on the CELPIP-LS Test

Band	Listening	Speaking
M	4.0%	2.0%
3	9.8%	4.5%
4	15.3%	12.6%
5	12.1%	20.9%
6	13.5%	28.9%
7	11.2%	14.1%
8	14.3%	11.4%
9	5.3%	2.6%
10	5.3%	1.3%
11	5.3%	1.1%
12	3.9%	0.6%

The above tables (Tables 4, 5a, and 5b) suggest that a larger proportion of test takers achieve lower bands on the CELPIP-General LS Test than on the CELPIP-General Test. It is important to note that, generally, individuals take the CELPIP-General and the CELPIP-General LS to meet different immigration requirements. The CELPIP-General Test is typically taken by the primary applicant for Canadian Permanent Residency. The CELPIP-General LS is required for applications for Canadian Citizenship. The minimum language proficiency requirements are currently different for permanent residency and citizenship applications. The different requirements may be a primary contributor to the observed difference in score profiles between the CELPIP-General and the CELPIP-General LS Tests.

Appendices A - E offer a more detailed breakdown of test taker performance by gender, first language, and country of citizenship.

4.2 Measurement Consistency

In statistics and psychometrics, measurement consistency is often referred to as reliability. It is an important component when evaluating the quality of a test. A reliable test gives us the same result repeatedly, assuming there is no change in the individual's ability. For example, if a test is designed to measure English language proficiency and has been administered to the same individual multiple times, the test scores should be approximately the same if the test taker's has not significantly improved their English proficiency during the period of the time. In contrast, an unreliable test produces inconsistent results every time, which greatly limits the value of the test scores.

Reliability could be estimated in a number of different ways. A higher value suggests a greater reliability of the test scores. A reliability of 0.80 and above is recommended according to the literature and industry standards. In this section, we report internal consistency for CELPIP Listening and Reading tests and rater agreement for CELPIP Speaking and Writing tests.

Internal consistency is a measure of whether test items intended to measure the same construct produce similar results. It is suitable for quantifying the reliability of tests that consist of many items, such as the Listening and Reading components on the CELPIP Test. Using Cronbach's alpha, the most common measure of internal consistency, we present the mean internal consistency estimates of all CELPIP Listening and Reading forms administered in 2019 (see Table 6).

Table 6: Reliability Estimates for the Listening and Reading forms on the CELPIP Test

Component	Average Reliability	Standard Deviation of Reliabilities
Listening	0.88	0.06
Reading	0.90	0.05

Table 6 shows that the mean internal consistency for the Listening and Reading forms was 0.88 and 0.90, respectively. These values suggest that, on average, the internal consistency within each of the CELPIP Listening and Reading forms are good. Additionally, the standard deviations among the Listening and Reading forms were small (0.06 for Listening and 0.05 for Reading), indicating good consistency between forms.

The reliability of the Speaking and Writing scores of the CELPIP Test is primarily maintained by assigning multiple raters to assess a test taker's performance and monitoring of rater agreement. The raters for the Writing and Speaking components of the CELPIP Test are highly proficient in English and are fully trained and certified by Paragon. Each of the test takers' writing tasks are rated independently by two accredited raters. If the scores awarded by the raters fall outside an acceptable tolerance of difference, the task is evaluated by a third rater. Each test taker's speaking performance is rated independently by three accredited raters. The responses are evaluated by two additional raters if the scores awarded by the original raters fall outside an acceptable tolerance of difference.

Since the evaluation process for the Writing and Speaking components relies on human judgements and the interpretation and application of a rating scale, variations in judgements are to be expected. Paragon constantly monitors rater agreement for quality control purposes. Table 7 shows the rater agreement for the Writing and Speaking components in 2019. Overall, the results indicate excellent consistency of judgement among raters.

Table 7: Rater Agreement for Writing and Speaking

	Speaking	Writing
Rater Agreement	88.2%	85.6%

5 Closing Remarks

The CELPIP Test continues to be an industry leader in language assessment. Testing is fully computer-delivered and available at over 70 locations across Canada and internationally, with test dates available every month. Paragon offers a range of test preparation materials, both for free and for purchase, to help you prepare for your test. Our CELPIP experts work in collaboration with official CELPIP Test Centers to provide the CELPIP Preparation Program to help test takers build the skills for each component of the CELPIP Test. For more information about test registration and preparation, visit the CELPIP website at https://www.celpip.ca/.

Contact Information

Paragon Testing Enterprises. 110-2925 Virtual Way, Vancouver BC V5M 4X5
research@paragontesting.ca

Appendices

The following appendices are provided as an overview of test scores based on different demographics.

Appendix A: CELPIP Scores by Purpose and Gender

Gender	Test Purpose	Listening Mean	Listening SD	Reading Mean	Reading SD	Writing Mean	Writing SD	Speaking Mean	Speaking SD
Female	IRCC	7.65	2.68	7.87	3.04	7.01	1.34	6.77	1.95
	NON-IRCC	8.63	2.13	8.59	2.51	7.39	1.09	7.71	2.05
Male	IRCC	7.83	2.67	8.42	2.96	7.08	1.38	6.76	1.99
	NON-IRCC	8.49	2.24	8.55	2.65	7.29	1.12	7.64	2.06

Appendix B: CELPIP-General Scores by Declared Nationality, Top 10 Represented Countries

Country	Listening Mean	Listening SD	Reading Mean	Reading SD	Writing Mean	Writing SD	Speaking Mean	Speaking SD
Brazil	8.74	2.27	9.12	2.57	7.19	1.02	6.79	1.25
Canada	9.12	1.99	9.08	2.44	7.61	1.00	8.47	2.08
China	7.78	2.33	8.13	2.65	6.85	1.03	6.14	1.15
India	8.15	2.14	8.17	2.56	7.12	1.07	6.74	1.25
Mexico	8.16	2.60	8.30	2.95	6.78	1.39	6.62	1.49
Nigeria	8.58	2.03	8.57	2.48	7.52	0.94	7.57	1.23
Philippines	5.65	1.96	5.20	2.33	5.93	1.15	5.44	1.01
South Korea	6.58	2.42	6.75	2.73	6.19	1.27	5.43	1.41
United Kingdom	10.46	1.61	10.94	1.72	8.43	1.06	10.59	1.46
United States	10.92	1.47	11.27	1.53	8.63	1.15	10.51	1.66

The countries are listed alphabetically

Appendix C: CELPIP-General Scores by Declared First Language, Top 10 Represented Languages

Language	Listening Mean	Listening SD	Reading Mean	Reading SD	Writing Mean	Writing SD	Speaking Mean	Speaking SD
Arabic	8.03	2.51	7.78	2.95	6.91	1.29	6.85	1.45
Chinese	7.84	2.31	8.14	2.64	6.87	1.03	6.20	1.18
English	9.65	2.18	9.87	2.56	7.97	1.24	9.17	2.12
Farsi	8.04	2.42	8.05	2.82	7.08	1.17	6.74	1.36
Hindi	8.39	2.09	8.42	2.49	7.23	1.03	6.92	1.20
Korean	6.64	2.43	6.80	2.74	6.22	1.26	5.46	1.41
Panjabi	7.42	2.12	7.18	2.51	6.73	1.10	6.39	1.23
Portuguese	8.71	2.29	9.07	2.61	7.16	1.08	6.78	1.28
Spanish	8.50	2.43	8.78	2.78	7.07	1.24	6.82	1.41
Tagalog	5.70	1.98	5.27	2.35	5.95	1.17	5.45	1.03

The language groups are listed alphabetically

Appendix D: CELPIP-LS Scores by Declared Nationality, Top 10 Represented Countries

Country	Listening		Speaking	
	Mean	SD	Mean	SD
China	5.89	2.31	5.29	1.27
India	6.11	2.46	5.73	1.56
Iran	6.10	2.35	5.79	1.34
Iraq	4.64	2.16	4.84	1.58
Israel	7.78	2.75	6.55	1.62
Mexico	7.38	2.47	6.35	1.40
Pakistan	5.87	2.33	5.80	1.54
Philippines	6.20	2.15	5.67	1.23
Syrian Arab Republic (Syria)	5.06	2.22	5.19	1.46
Ukraine	6.82	2.34	5.98	1.39

The countries are listed alphabetically

Appendix E: CELPIP-LS Scores by Declared First Language, Top 10 Represented Languages

Language	Listening		Speaking	
	Mean	SD	Mean	SD
Arabic	5.47	2.44	5.39	1.55
Chinese	6.04	2.33	5.36	1.26
English	7.46	2.71	7.52	2.39
Farsi	5.90	2.33	5.71	1.38
Panjabi	5.05	1.99	5.10	1.35
Portuguese	8.35	2.37	6.86	1.41
Russian	7.12	2.55	6.12	1.45
Spanish	7.39	2.53	6.30	1.48
Tagalog	6.25	2.10	5.70	1.22
Urdu	5.92	2.30	5.83	1.48

The language groups are listed alphabetically