

CELPIP Annual Report of 2020 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 assessment of general English language proficiency. It is administered by Paragon Testing Enterprises Inc.

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, informative, and defensible. Paragon works closely with test centres to make certain that the CELPIP Test is administered in a way that is fair, secure, and accessible to all individuals who wish to take the exam.

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	47 - 55 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.	Min 38
Reading	55 - 60 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.	Min 38
Writing	53 - 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

1.3 Scoring and Reporting of Test Results

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

Test takers receive a score report that provides a score for each component. The multiple-choice items in the Listening and Reading Components are scored by computer. 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.

CELPIP component scores are reported in 11 bands: M, 3 - 12. 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

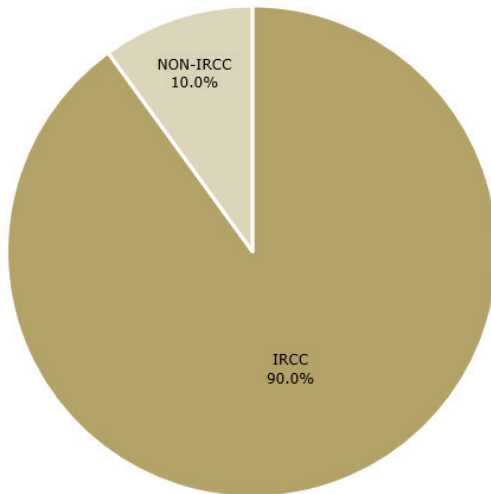
CELPIP Level	CELPIP Descriptor	CLB Level
NA	Not Administered: test taker did not receive this test component	/
M	Minimal proficiency or insufficient information to assess	1, 2
3	Some proficiency in limited contexts	3
4	Adequate proficiency for daily life activities	4
5	Acquiring proficiency in workplace 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 daily lives, study, or work. There are, inevitably, small mismatches between the tasks that test takers complete as part of the test and the tasks that they need 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 results.

2 Test-Taking Population

This section presents an overview of the test takers who took the CELPIP Test in 2020. Figure 1 presents the distributions of test takers' self-reported purpose for taking the test. The most commonly reported purpose for taking the CELPIP Test is for Canadian immigration and citizenship applications (IRCC = 90.0%).

Figure 1: Distribution of CELPIP Test Takers by Purpose for Taking the Test



Note: IRCC=Immigration, Refugees and Citizenship Canada

Figure 2 shows the distribution of test takers for the CELPIP-General and the CELPIP-General LS Tests in the total test-taker population.

Figure 2: Distribution of CELPIP-General and CELPIP-General LS Test Takers

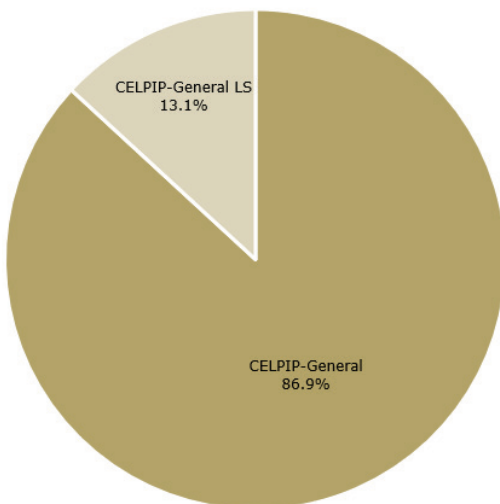


Table 3 shows that the majority of the CELPIP-General and CELPIP-General LS test takers were between 20 and 40 years old (81.38%). A comparison of distributions reveals that the CELPIP-General Test has a younger demographic, with a larger group between 20 and 30, while the CELPIP-General LS test taker population has a slightly larger group between ages 31 and 55.

Table 3: Distribution of CELPIP Test Takers by Age (%)

Age Group	% of Total Test Population	% of CELPIP-General	% of CELPIP-General LS
< 20	0.2	0.2	0.2
20-25	14.9	16.5	4.8
26-30	28.7	30.9	13.9
31-35	21.7	21.3	24.2
36-40	16.1	14.8	25.0
41-45	9.6	8.5	17.2
46-50	5.4	4.7	10.3
51-55	2.4	2.1	4.2
> 55	0.8	0.9	0.2

3 Test Statistics

3.1 Score Descriptions

Tables 4 presents the mean and standard deviation for the CELPIP-General and CELPIP-General LS component scores. The mean score is the simple average of all test takers' scores, and the standard deviation quantifies how scores are spread out from the mean.

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

Test	Component	Mean	Standard Deviation
CELPIP-General	Listening	8.21	2.46
	Reading	8.28	2.89
	Writing	8.02	1.93
	Speaking	7.83	2.21
CELPIP-General LS	Listening	6.56	2.53
	Speaking	6.57	2.02

Tables 5a and 5b show the percentage breakdown of test takers who received each component band score on the CELPIP-General and CELPIP-General LS Tests.

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

Band	Listening	Reading	Writing	Speaking
M	0.6	2.7	0.2	0.3
3	2.2	2.9	0.7	0.8
4	5.5	6.0	1.7	3.1
5	6.7	7.3	5.3	8.3
6	10.0	10.7	11.0	19.0
7	11.8	10.2	21.3	16.7
8	20.8	11.0	27.5	20.2
9	8.7	12.4	10.5	9.6
10	10.7	7.5	9.2	5.9
11	12.0	7.4	5.8	7.1
12	10.9	21.8	6.6	9.2

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

Band	Listening	Speaking
M	3.1	1.9
3	7.9	3.6
4	14.4	8.1
5	12.0	15.5
6	14.0	24.4
7	12.3	15.4
8	15.7	16.2
9	6.0	6.9
10	6.0	3.3
11	5.1	2.9
12	3.5	1.8

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, which may be a primary contributor to the observed score profile differences between the CELPIP-General and the CELPIP-General LS Tests.

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

3.2 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 consistently, assuming 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 has not significantly improved their English proficiency during the period of time. In contrast, an unreliable test produces inconsistent results each time, which greatly limits the value of the test scores.

Reliability can be estimated in multiple ways. In general, 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 (e.g., George & Mallery, 2016). In this section, we report internal consistency for CELPIP Listening and Reading tests and rater agreement for CELPIP Speaking and Writing tests.

3.2.1 Internal Consistency for the Listening and Reading Components

Internal consistency is a measure of whether test items designed 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. To measure internal consistency, Cronbach's alpha was calculated for each CELPIP Listening and Reading form administered in 2020 (see Table 6).

Table 6: Reliability Estimates for CELPIP Listening and Reading forms

Component	Average Reliability	Standard Deviation of Reliabilities
Listening	0.89	0.02
Reading	0.90	0.01

Table 6 shows that the mean internal consistency values for the Listening and Reading forms were 0.89 and 0.90, respectively. These values suggest that, on average, there was high internal consistency within each of the CELPIP Listening and Reading forms. Additionally, the mean distances from the median were small (0.02 for Listening and 0.01 for Reading), indicating the forms performed similarly in terms of their internal consistency.

3.2.2 Rater Agreement for the Writing and Speaking Components

The reliability of the Writing and Speaking scores of the CELPIP Test is maintained operationally by assigning multiple raters to assess a test taker's performance and regular 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 writing task is rated independently by two accredited raters. If the scores awarded by the raters disagree, the task is evaluated by a third rater. Each test-taker's speaking performance is rated independently by three accredited raters. Two additional raters will evaluate the responses if the scores awarded by the original raters disagree.

Since the evaluation process for the Writing and Speaking components relies on human judgment and the interpretation and application of a rating scale, variations in judgments 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 2020. Overall, the results indicate excellent consistency of judgment among raters.

Table 7: CELPIP Rater Agreement for Writing and Speaking (%)

	Speaking	Writing
Rater Agreement	90.7	90.5

4 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. Our CELPIP experts work in collaboration with official CELPIP Test Centres to provide the CELPIP Preparation Program to help test takers build the skills for each CELPIP Test component. For more information about test registration and preparation, visit the CELPIP website at <https://www.celpip.ca/>.

Reference

George, D., & Mallery, P. (2016). *IBM SPSS statistics 23 step by step: A simple guide and reference*. Routledge.

Appendices

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

Appendix A: CELPIP Scores by Test Purpose

Test Purpose	Listening		Reading		Writing		Speaking	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
IRCC	7.92	2.56	8.23	2.93	7.96	1.95	7.54	2.19
NON-IRCC	8.73	2.13	8.73	2.54	8.46	1.72	8.75	2.21

Appendix B: CELPIP-General Scores by the Top 10 Declared Nationality

Country	Listening		Reading		Writing		Speaking	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
India	7.95	2.17	7.87	2.63	7.74	1.59	7.37	1.61
China	7.70	2.41	8.05	2.72	7.47	1.49	6.72	1.48
Philippines	5.99	2.05	5.52	2.50	6.59	1.57	6.00	1.38
Canada	9.12	1.96	9.15	2.37	8.82	1.66	9.51	2.08
Brazil	8.78	2.25	9.18	2.56	8.05	1.55	7.57	1.56
South Korea	6.72	2.32	6.84	2.65	6.75	1.50	5.92	1.62
United States	10.94	1.41	11.40	1.39	10.65	1.47	11.26	1.23
United Kingdom	10.38	1.69	10.86	1.81	10.24	1.58	11.20	1.33
Nigeria	8.63	2.08	8.78	2.54	8.76	1.67	8.71	1.62
Mexico	8.11	2.41	8.25	2.83	7.51	1.80	7.39	1.75

Appendix C: CELPIP-General Scores by the Top 10 Declared First Languages

Language	Listening		Reading		Writing		Speaking	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
English	9.69	2.10	9.92	2.46	9.53	1.88	10.12	2.00
Panjabi	7.22	2.11	6.89	2.51	7.16	1.43	6.90	1.55
Chinese	7.72	2.41	8.04	2.74	7.50	1.52	6.76	1.52
Hindi	8.29	2.07	8.31	2.55	7.97	1.51	7.70	1.55
Spanish	8.49	2.34	8.73	2.70	7.90	1.75	7.62	1.74
Tagalog	6.11	2.09	5.69	2.55	6.68	1.59	6.09	1.43
Portuguese	8.67	2.32	9.04	2.67	7.94	1.69	7.51	1.66
Korean	6.75	2.32	6.86	2.64	6.76	1.50	5.94	1.62
Arabic	7.89	2.43	7.64	2.92	7.66	1.89	7.66	1.87
Farsi	8.11	2.26	8.00	2.66	7.83	1.62	7.47	1.60

Appendix D: CELPIP-LS Scores by the Top 10 Declared Nationalities

Country	Listening		Speaking	
	Mean	SD	Mean	SD
India	6.45	2.42	6.40	1.91
Iran	6.28	2.23	6.29	1.56
Syrian Arab Republic (Syria)	5.65	2.41	5.89	1.83
Pakistan	6.08	2.14	6.35	1.85
Mexico	7.52	2.28	7.25	1.82
China	6.03	2.21	5.90	1.39
Philippines	6.57	2.19	6.37	1.63
Iraq	5.14	2.20	5.69	1.91
Ukraine	6.77	2.45	6.55	1.71
Brazil	8.29	2.24	7.63	1.61

Appendix E: CELPIP-LS Scores by the Top 10 Declared First Languages

Language	Listening		Speaking	
	Mean	SD	Mean	SD
Arabic	5.84	2.46	6.09	1.91
Spanish	7.47	2.35	7.10	1.83
Farsi	5.99	2.26	6.18	1.59
English	7.64	2.52	8.35	2.39
Urdu	6.16	2.13	6.40	1.87
Russian	7.30	2.57	6.74	1.80
Chinese	6.13	2.23	5.93	1.43
Panjabi	5.54	2.22	5.66	1.80
Portuguese	8.14	2.29	7.52	1.75
Tagalog	6.60	2.18	6.35	1.55