CELBAN

CANADIAN ENGLISH LANGUAGE
BENCHMARK ASSESSMENT FOR NURSES

FACTS & FIGURES

Issue 3

CELBAN SPEAKING TEST RENEWAL

In December 2015, a survey was sent to all IENs who had completed CELBAN through The CELBAN Centre. Of 166 IENs who responded, **86 per cent reported that they feel CELBAN benefitted them** in terms of learning about their nursing communication skills. Many reported that the unofficial test results helped them learn more about the language skills they need to work on in the health care field, and nursing in particular.

Over 10 years ago, CCLB embarked on a multi-phase project to develop CELBAN in response to a recognized need by nursing profession stakeholders for a specialized English language assessment tool to evaluate the English language communication skills of IENs seeking registration in Canada.

By 2014, CELBAN had established itself as a preferred option for many IENs. As an occupation-specific communication assessment, CELBAN provides content that is meaningful, interesting and relevant to prospective nursing professionals.

With the launch of The CELBAN Centre in 2014, Touchstone Institute initiated a test renewal project to evaluate and update the existing CELBAN content and procedures after a decade of heavy usage. The CELBAN scoring grid was also due for a review. Original CELBAN scoring criteria was based on descriptors from the CLB 2000; since then, a pan-Canadian revision process had produced a newer (2012) version of the CLB. An important part of the renewal project would thus involve revising the grid to align with descriptors of criteria as expressed in the 2012 CLB document.

This issue of CELBAN Facts & Figures presents phase 1 of the project: a review and redevelopment of the CELBAN Speaking Test.

CELBAN Speaking Test

The CELBAN Speaking Test features eight tasks that engage candidates in discussions and role plays. The discussion tasks elicit function-specific discourse on health-related topics, and the role-play tasks prompt interactions typical of the target professional context. The test takes between 20 and 30 minutes and is facilitated and scored by two trained assessors.

The two-assessor model allows one assessor to facilitate and interact with the candidate while the other observes and evaluates. The assessors switch between facilitating and observing so that both are fully involved throughout the assessment. When the test is finished, the two assessors confer and compare notes to arrive at a final score. Scores are assigned in reference to a grid based on criteria drawn from the Canadian Language Benchmarks (CLB). If the two assessors cannot reach consensus, a third rating is sought.

UNIQUE FEATURES OF CELBAN

The CELBAN Speaking Test identifies multiple benchmarks for the purpose of providing diagnostic and constructive feedback to candidates. This unique feature of the test is a strength that sets it apart from other standardized assessments. The quality of the feedback is continuously monitored for its relevance and positive washback, so that ongoing improvements can be made.

Process for Renewal

STEP 1: CELBAN Speaking Test Review

The review of the CELBAN Speaking Test began with an analysis of test content and procedures from a current language proficiency testing perspective.

FINDINGS:

TEST PROCEDURES ▶ Test protocols are direct and robust; tasks provide a suitable range of opportunities for expression using varied language functions, structures and vocabulary.

TEST CONTENT ▶ New content needs to be developed to replace test materials that have been exposed over 10 years of administration.

TEST ASSESSOR MODEL ▶ Double assessor model provides an opportunity to discuss and debate a candidate's strengths and weaknesses and determine a defensible result.

SCORING ▶ Speaking grid requires updating and re-alignment.

The results of the Speaking Test review helped us map a draft process for renewal that would build on CELBAN's existing strengths while continuing to investigate possibilities for improvement, enhancement and refinement.

STEP 2: Assessor Feedback

Assessor feedback was collected in two forms:

1) An online survey of trained assessors gathered feedback about test content and procedures.

2) A small focus group of experienced assessors led to in-depth discussion about their impressions of the test content, as well as scoring procedures and grid.

GENERAL FINDINGS:

Assessors affirmed that the test provided comprehensive domain coverage; provided ample time and opportunity to elicit, observe and rate language performance; and offered test takers a meaningful and relevant assessment experience.

RECOMMENDATIONS:

• Update content of some test tasks for increased effectiveness and appropriateness.

• Retain the original formula used to define the final benchmarks.

• Reduce the number of benchmark levels represented on the scoring grid – four benchmarks give assessors enough information to make valid decisions.

• Update individual criteria in the grid to reflect 2012 CLB levels.

STEP 3: Subject Matter Experts (SME) Consultations

A focus group of nursing consultants convened to act as SMEs. They reviewed each role play and discussed the quality and relevance of the current tasks, then explored alternative scenarios that would fit the same genre, language function and specifications. These SME consultations helped confirm new nursing-related content.

STEP 4: Development of Renewed Forms

A new Speaking Test protocol was designed around four key tasks that had been identified as crucial to the assessment. From this protocol, new test content was developed for each task. A new scoring grid was also produced, based on the feedback from consultations with assessors and nursing SMEs.

SUMMARY OF TEST PROTOCOL CHANGES:

• Labelling and sequencing of the protocol was changed to enhance the flow of the assessment. Tasks are now simply numbered as Task 1, Task 2, Task 3, etc.

• Instructions were revised to be more simple, precise and concise.

• Terminology was reviewed and updated for accuracy.

• A probing task was reworked into a role-play between a nurse and a colleague, which better suited the context and introduced peer dynamics. Nurse to nurse communication was previously missing from the test.

• Some tasks were reordered to be more efficient and logical for both the test taker and the assessors.

• The intake form used for a role play task was abbreviated to focus test takers on constructing a good variety of initial and follow-up questions without straying off task.

SUMMARY OF SCORING GRID CHANGES:

• The number of benchmarks was reduced from six to four.

• Some criteria were conflated and renamed for clarity.

• Descriptors were reworded to align with language in the 2012 CLB document.

• Criteria were reordered.

STEP 5: Refinement Workshop

A workshop, facilitated by the test development lead, was conducted with four CELBAN assessors. The purpose of the workshop was to observe the assessment flow in action to review and refine the protocol, tasks and scoring prior to the pilot test. Assessors represented a mix of experience – two were new to CELBAN and two were highly experienced with the test – to ensure the materials would ultimately be appropriate and accessible for all assessors. Assessors were oriented to the new material and then participated in a series of four trial assessments, each with a different volunteer CELBAN candidate. After each trial, assessors discussed the effectiveness of the materials and made suggestions for adjustments.

STEP 6: Assessor Feedback

The pilot test was conducted with four experienced assessors, representing different parts of the country, and 46 pilot subjects. Assessors worked together in pairs, taking turns facilitating and observing candidates.

Live pilot administrations were recorded and submitted to a second review by the alternate assessor pair. The procedure for video assessment was structured to mirror live assessment as closely as possible.

LIVE ASSESSMENT
1. Assessor pairs participated in live assessment with a volunteer CELBAN candidate.
2. After the candidate exited the exam room, each assessor completed the scoring grid and calculated the corresponding benchmark independently.
3. Assessors compared their independently assigned scores and discussed the candidate's strengths and weaknesses.
4. After considering the overall performance, the two assessors agreed on a final benchmark.

VIDEO ASSESSMENT
1. Individual assessors watched a set of four recorded assessments at separate locations.
2. After viewing the recording, each assessor completed the scoring grid and calculated the corresponding benchmark independently.
3. Immediately following the viewing, assessor pairs participated in a telephone conference to discuss the candidate's strengths and weaknesses.
4. After considering the candidate's performance, the two assessors agreed on a final benchmark.

STEP 7: Reliability Check

The reliability check compared the raw scores calculated from each assessor's independent completion of the scoring grid. The pilot study was designed to make the most of independently assigned scores, and offers us 276 points of comparison.

Inter-rater Score Comparisons

Comparing all the scores assigned independently by the four assessors shows that there was excellent agreement across all assessor pairings. There are no differences greater than one benchmark, and the average margin of agreement for each pair is identical at 0.3, or less than one third of a benchmark.

Differences Between Scores for All Assessor Pairings

Assessor Pairing	Number of Cases in which Scores of the Two Assessors Differ by ...				Total Pairings	Average Difference
	0 – 0.2	0.3 – 0.5	0.6 – 0.8	0.9 – 1.00		
AB	26	13	4	3	46	0.3
AC	23	15	4	4	46	0.3
AD	25	15	4	2	46	0.3
BC	28	9	8	1	46	0.3
BD	19	18	7	2	46	0.3
CD	24	11	6	5	46	0.3
TOTAL	145	81	33	17	276	---
%	53%	29%	12%	6%	100%	---

Live versus Video Scores

The comparison of independently assigned raw scores indicates a high degree of consistency between assessors – both in the live pilot environment and the remote video assessment. On average, assessors differed in their independently assigned ratings by one third of a benchmark or less, and in a high proportion of cases, they attain complete agreement. This occurs regardless of whether an assessor administers the assessment live or scores remotely by viewing the test on video.

Differences between Scores of Two Live Assessors

Score Difference	Number of Occurrences	Percentage of Occurrences
0.9	1	2%
0.8	1	2%
0.7	0	0%
0.6	0	0%
0.5	1	2%
0.4	6	13%
0.3	6	13%
0.2	7	15%
0.1	14	30%
0	10	23%
Total	46	100%

Differences between Scores of Two Video Assessors

Score Difference	Number of Occurrences	Percentage of Occurrences
0.9	0	0%
0.8	0	0%
0.7	0	0%
0.6	0	0%
0.5	5	11%
0.4	4	9%
0.3	7	15%
0.2	10	22%
0.1	9	19%
0	11	24%
Total	46	100%

Video ratings are not normally a part of CELBAN; however, the pilot allowed us an opportunity to observe the relationship between final rounded benchmarks of the pairs of assessors in both live and video contexts. As the following table illustrates, only small discrepancies (in bold) were evident. Viewing assessors differed from live assessors in their final benchmark decision in 37 percent of cases. This could be attributed to the experiential difference in actively versus passively interacting with test takers, as well as the different processes involved in live versus video assessment. In the video assessment, discussions between the paired assessors occurred after they had viewed a series of four recorded assessments. In the live pilot administration, assessors discussed the candidate's performance immediately following the assessment.

Average Score by Assessor

Live Pairing	Video Pairing	Average Difference
AB	CD	0
AC	BD	0
AD	BC	**0.1**
BC	AD	0
BD	AC	0
CD	AB	**0.2**

Reliability Considerations

CELBAN is a CLB-based assessment, which means that the final result is not necessarily a whole number. There are no decimal points or half bands in the benchmark system, and CELBAN therefore relies on a rounding formula for the determination of final scores. Results of .8 and higher are rounded up, and those of .7 and lower are rounded down. When it comes to final scores, incremental degrees of agreement do not exist. The inescapable reality of CLB-based assessment is the fact that the smallest possible discrepancy for a final result is always a full benchmark.

A candidate's final rounded benchmark score plays a critical role in his or her journey to professional practice. The minimum score accepted by Canadian nursing regulators for the CELBAN Speaking Test is 8. In order to ensure score reliability, The CELBAN Centre's Assessor Training Program places heightened emphasis on the distinction between benchmark levels 7 and 8. Trainees receive rigorous coaching on making decisions at this critical threshold and have opportunities to practise doing so consistently.

The CELBAN Centre administered 1,985 CELBAN tests across Canada in 2015. As of June 2016, we have already administered over 1,000 tests and expect registration to exceed 2,000 for the year.

SUMMARY

Renewal of the CELBAN Speaking Test has provided an opportunity to review and clarify procedures, update test content and evaluate the test's reliability with the input of new and experienced assessors and nursing subject matter experts.

As CELBAN moves forward, Touchstone Institute and the Centre for Canadian Language Benchmarks remain committed to a plan of responsible test maintenance that involves ongoing development, quality control checks, gathering of statistical data, and research studies that add to the existing body of validity evidence to ensure that the test remains suited to its intended purpose.

This issue of *CELBAN Facts & Figures* is based on the CELBAN Speaking Test Renewal Report, written by Gail Stewart for Touchstone Institute.

Building Capacity, Ensuring Quality

CELBAN Assessors

In just under two years, The CELBAN Centre has trained 53 new speaking assessors across Canada, essentially doubling our assessor pool. In doing so, we've enabled test sites to increase their test session capacities to accommodate more test-takers and meet the demands of IENs looking for proof of language proficiency.

CELBAN assessors are carefully selected through a process of PLAR (Prior Learning Assessment and Recognition) to ensure that they bring relevant, appropriate and adequate experience to their roles. The CELBAN assessor community, now at 90, has an invaluable wealth of knowledge in the world of language assessment, occupation-specific language training and cross-cultural communication.

A Quality Framework

The CELBAN Centre has implemented a quality assurance framework to maintain the consistency of CELBAN speaking scores, inter-rater reliability and the integrity of the testing experience. Members of the CELBAN quality assurance team conduct periodic spot checks on speaking test recordings to provide assessors with feedback on the characteristics of their testing interactions, as well as review paperwork to validate scoring decisions and test-taker feedback.

Findings from these checks will help to guide planning for annual calibration activities, as well as bi-annual assessor recertification. The CELBAN Centre has included input from experienced assessors and in-house staff who work together to facilitate spot checks, develop criteria for feedback and plan for future training, recertification and calibration sessions.

Assessor Community Platform

As of summer 2016, The CELBAN Centre will activate an assessor community platform using Tutela to conduct scoring calibration exercises, standardize test interactions and as a forum for assessors to share resources and experiences.

In preparation for the launch of the new Speaking Test versions, all CELBAN assessors will undergo a recertification and new versions training in September 2016. To facilitate this process, The CELBAN Centre will develop a training webinar hosted by Tutela. Training will include instruction on navigating new speaking tasks, with an emphasis on eliciting target language use, as well as adapting to the updated scoring criteria grid.

CELBAN Facts & Figures is a biannual report series is prepared by the Communication Competency Program at Touchstone Institute.

TOUCHSTONE
INSTITUTE
COMPETENCY EVALUATION EXPERTS

The CELBAN Centre is the national administrative centre for the Canadian English Language Benchmark Assessment for Nurses (CELBAN).

THE
CELBAN CENTRE