CELBAN

CANADIAN ENGLISH LANGUAGE
BENCHMARK ASSESSMENT FOR NURSES

FACTS & FIGURES
ISSUE 5

CELBAN LISTENING TEST RENEWAL

Since 2014, Touchstone Institute has been the national test administrator responsible for test delivery and quality assurance. The Facts & Figures series documents and makes transparent the CELBAN test renewal process undertaken from 2014-2018.

Background

CELBAN is built on communication that is essential to nursing and represents authentic language use. The test is task-based and embeds the use of language within a profession-specific framework with a focus on contexts that are fundamental to nursing practice. There is a separate test instrument for each of the four language skills - listening, speaking, reading and writing. CELBAN test results are referenced to the Canadian Language Benchmarks (CLB), a descriptive scale of communicative ability in English as a Second Language (ESL).

The CELBAN test renewal project was launched in 2014, when the CELBAN Centre was first established at Touchstone Institute. As the original CELBAN test model had been in place for several years, the purpose of the renewal project was to develop additional forms of the test that would retain the best features of the original model, while also introducing some new task types and fresh content.

The renewal process involved a review of content and format, development of new tasks and items, pilot testing, data analysis, construction of renewed forms and a launch of the renewed procedures.

REVIEW OF TEST CONTENT AND FORMAT → DEVELOPMENT OF NEW CONTENT → PILOT TESTING OF NEW CONTENT → DATA ANALYSIS & CONSTRUCTION OF FORMS → LAUNCH OF THE RENEWED TEST
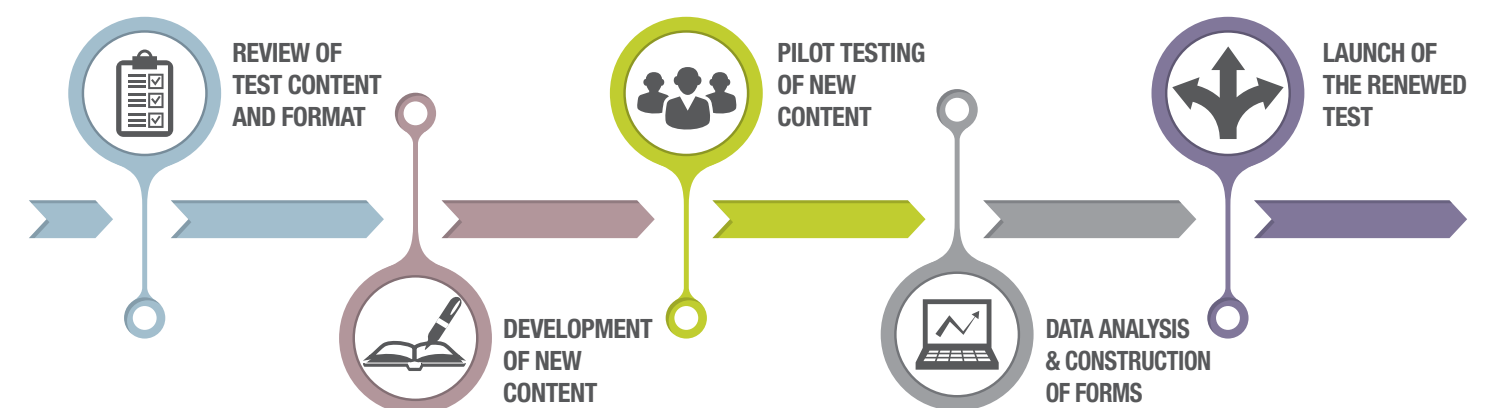
This issue of *CELBAN Facts & Figures* presents a summary of the work that was carried out to produce renewed forms of the CELBAN Listening Test. For information on the Speaking Test renewal, see CELBAN Facts & Figures, Issue 3, and for the Reading Test, see Issue 4.

Review and Consultations

The original CELBAN Listening Test consisted of video and audio scenarios, supporting a total of approximately 75 multiple-choice items. The first step in the renewal process was a review of all of the original test content by a test development specialist. Each scenario was evaluated, and all of the audio and video content was reviewed for clarity, relevance and production quality. The items in the test were scrutinized for quality, independence and integrity, and they were classified based on the listening abilities and strategies required to respond correctly.

Findings from this research confirmed the communicative strengths of the test model. The nursing scenarios were found to be authentic and representative of relevant situations and contexts. Some of the video scenarios were quite lengthy, which meant that large numbers of test items were tied to a single conversation and context. This aspect of the test was noted for further investigation.

Following this review, consultations were organized with a small group of nurses serving as subject-matter experts (SMEs). These experts were provided with background information and test content for review. They were then brought together in a focus group to share impressions and provide feedback. In general, the SMEs had positive impressions of the test. They thought that the approach was engaging, and they found the content to be authentic and relevant. They observed that some of the scenarios were rather long, and they suggested that shorter scenarios might be more effective, as this would allow a larger number of video and audio clips to be included in each test form, thereby representing a greater variety of situations and contexts.

The consultation process also took into account administration procedures. The original CELBAN Listening Test consisted of digital files viewed on a large screen and two booklets that were given to candidates. One booklet contained the instructions and test questions, while the other was used to record responses. This meant that candidates had to pay attention to the video screen, while also listening to scenarios and working between two booklets. During the consultations it was noted that this approach created a significant cognitive load that was not related to the listening functions being assessed.

Revisions to the Design and Model

Based on the expert reviews and consultations, it was determined that the types of interactions included in the original Listening Test should be retained and that these should be supplemented with additional content to more broadly address the range of situations and contexts that nurses encounter. The renewed test model is therefore based on shorter video and audio scenarios, which provide greater flexibility for test development. Shorter clips can be mixed and matched to create alternate forms of the test, and a larger number of scenarios can be included in each test form, providing fuller coverage of the listening domain with a broader range of situations, speakers and scenarios. This broadened coverage helps to ensure that test takers are not disadvantaged by having too many items linked to a single context or situation. It also reduces the likelihood that listeners will become fatigued or overwhelmed by the length of the clips.

Administration procedures were also adjusted so that candidates would no longer have to work with two separate booklets while simultaneously viewing a video screen and listening to conversations. The administration has been simplified so that there is only one booklet, which includes all of the instructions and test items. Candidates make their responses directly in that booklet.

Content Development and Production

A very small and highly experienced development team was established to work on content for the renewed Listening Test. These script and item writers worked closely with a nurse advisor, who provided feedback on the accuracy and authenticity of the content. All content was reviewed and revised before the production process began.

The production process was lengthy and complex. It began with the selection of actors and voice professionals who had the talent, delivery and clarity of articulation required to perform the scenarios. A professional audio-visual production company was engaged to scout locations, set up a shooting schedule and provide the necessary props, set design pieces and wardrobe. For the video scenarios, actors were brought in for a day of rehearsal prior to shooting their scenes. The shooting was carried out over a five-day period, under the guidance of a professional director, camera crew, language testing expert and nurse advisor. Audio recording took place in a professional recording studio, supervised by an experienced sound engineer.

The video shoots and audio recording sessions were followed by a post-production process that involved reviewing hours of footage, selecting the best takes, editing video and audio clips, designing a visual framework, assembling, reviewing and finalizing digital content and preparing pilot test forms. Paper-based test booklets were designed and produced so that pilot test candidates would be able to choose their responses while viewing the digital files on a large screen.

Pilot Testing and Data Analysis

Two pilot test forms, X and Y, were constructed using the content that had been developed, and 178 candidates were recruited to pilot the test items. As is typical in many occupation-specific test development projects, the practicalities of recruitment and administration imposed restrictions on the pilot sample size. Because the sample was not large, care was taken to match the characteristics to the population. Participants were recruited across the range of ability levels typical of actual CELBAN test takers, and the sample was carefully stratified. Each candidate attempted all of the items in both pilot test forms, with the order of administration counterbalanced so that half of the candidates at each ability level took Form X followed by Form Y, and the other half took Form Y followed by Form X.

The pilot testing was co-ordinated and administered by the CELBAN management team from Touchstone Institute. To ensure that conditions were standardized, each session was proctored according to a scripted procedure. The pilot testing ran smoothly, and the process was not interrupted or affected by any unusual circumstances or unforeseen events.

For data analysis purposes, the items in the pilot were treated as one long test. Candidates' overall scores on the full item set were used to rank ability within the group. The psychometric team from Touchstone Institute calculated difficulty and discrimination indices for all of the items, and these values were used to eliminate poorly functioning items and to select an appropriate range of content for the operational test.

Item Calibration

Following the data analysis, test items were referenced to the Canadian Language Benchmarks (CLB) by a panel of experts.

The steps in this calibration process were as follows:

1. Independently, each panelist attempted all of the items in the pilot test, making notes on the language abilities needed to answer each question and the level of difficulty of each item according to the benchmarks scale.

2. Each panelist received a chart with the items ranked in order from the most to the least difficult based on the data analysis. Independently, each expert grouped the items into CLB levels, using the listening descriptors in the CLB 2012 document.

3. The panel was brought together to discuss the levels they had assigned and to reach agreement on the item groupings and levels.

Summary

The original CELBAN Listening Test model has been updated, and renewed forms of the test have been configured based on the results of pilot testing, data analysis and item calibration. Each renewed form contains 78 test items. The new listening scenarios are relevant to current nursing policies and practices, and the video and audio clips represent a range of nursing contexts, situations and professional tasks. Conversations include nurses, doctors, patients, family members and consulting professionals.

The test items address a broad range of listening skills and strategies at appropriate levels of difficulty to provide candidates with opportunities to demonstrate different facets of their listening ability. Specificity to the nursing domain has been accomplished by a team of developers working closely with nurse experts. Care has been taken to ensure that the ability to respond to test items is based on language proficiency and not on clinical knowledge.

Under the stewardship of Touchstone Institute, with input from language specialists and consulting nurses, the CELBAN content and procedures have evolved to reflect the current nursing context. The renewed model preserves the best features of the original test and builds on the solid foundation that was established in the CCLB CELBAN development project. Calculations indicate that the renewed test has a high degree of reliability, with a KR20 coefficient of .91 for internal consistency.

In the coming years, research and development will continue, with ongoing test maintenance and quality control. The Touchstone Institute psychometric team will continue to conduct regular data analysis, in keeping with Touchstone Institute's commitment to ensuring reliability, fairness and integrity in assessment.

Facts & Figures Reports are developed and published by The CELBAN Centre to support transparency in test research and development and document our experiences and activities as administrators of CELBAN. The publication is available from The CELBAN Centre's website. For more information, contact celban@tsin.ca.

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

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

TOUCHSTONE
INSTITUTE
COMPETENCY EVALUATION EXPERTS

THE
CELBAN CENTRE