Issued by:
The Canadian Council for Practical Nurse Regulators (CCPNR)
on behalf of Canadian Nurse Regulators Collaborative

Proposal submission deadline:
12:00 EDT August 17, 2018

REQUEST FOR PROPOSAL:

Language Proficiency Tests and Benchmarks for Licensure to Practice as a Nurse in Canada

July 27, 2018

TABLE OF CONTENTS

SCOPE

The Canadian Council for Practical Nurse Regulators (CCPNR) on behalf of the Canadian Nurse Regulators Collaborative (CNRC) is seeking a subject matter expert to complete an environmental scan of the fluency requirements appropriate for safe nursing practice in Canada. The objectives below form the key deliverables for the RFP.

Objectives

To:

- identify trends and their impact in language proficiency testing for safe nursing practice in Canada;
- review and assess the benchmarks for the accepted English fluency/proficiency tests: Official CELBAN and Academic IELTS, and accepted fluency/proficiency French test: TEF for Registered Nurses (RN) and Licensed/Registered Practical Nurses (LPN) resulting from the initial national standards setting exercise in 2010;
- determine enhancements necessary to the 2010 benchmark scores including the applicability of these scores to Registered Psychiatric Nurses (RPN) in Canada; and
- propose directions and next steps for consideration by CNRC.

Deliverable

Report outlining the:

- trends and impact in language proficiency testing;
- results of benchmark assessment; and
- proposed directions and next steps for consideration.

Building on the work completed for the 2010 performance standards setting exercise, the research will combine primary and secondary methods to achieve the objectives of the scan. This will include document and literature review, as well as consultations with stakeholders in nursing regulation and language proficiency testing.

Prior to commencing the research, the successful respondent will outline the work plan providing details of the activities, outputs, responsibilities, and timelines for steering committee approval.

BACKGROUND and CONTEXT

Recognizing the need for improved collaboration and communication between nurse regulators, the chairs of the Canadian Council for Practical Nurse Regulators (CCPNR), Canadian Council of Registered Nurse Regulators (CCRNR) and the Registered Psychiatric Nurse Regulators of Canada (RPNRC) established the CNRC and hosted the first ever forum for nurse regulation in Canada on July 6, 2017. CNRC is established to provide a forum for nurse regulators to collaborate and exchange knowledge on common issues and trends.

The need to review English as a Second Language benchmarks was identified as a common issue for all nurse regulators to collaborate and move forward. By way of background, a National Fluency Working Group (NFWG) was established in 2008 to identify the fluency requirements appropriate for safe nursing practice and to explore whether there was national agreement to harmonize the requirements. The group consisted of representatives of nursing regulators from Registered Nurses, Licensed/Registered Practical Nurses and Nurse Practitioners across Canada. RPNs were represented at various stages of the work. Several recommendations were proposed including acceptable language proficiency tests and passing scores for English and French language proficiency for both Canadian and international applicants. Appendix A provides the 2010 draft report *Setting performance standards on three English language proficiency tests for licensure to practice as a nurse in Canada*.

REPORTING

The successful respondent will work closely with the committee's chairperson and report to, and collaborate with, a steering committee consisting of representatives of Registered Nurses, Licensed/Practical Nurses and Registered Psychiatric Nurses.

The steering committee will approve all final deliverables in principle and recommend next steps and/or recommendations for consideration to the CNRC chairs. CNRC chairs and their respective boards will have final approval of the deliverables. The role and responsibility of the steering committee is to oversee the project, provide support as required and recommend next steps for the CNRC chairs' consideration.

BUDGET and TIMELINE

The total budget for the environmental scan is not to exceed $10,000.00. This total amount includes all professional and administrative fees and HST. Teleconference or webinar fees are separate from the total budget.

The project will commence upon contract signing. A final environmental scan report approved in principle by the steering committee is expected by *December 31, 2018*.

PROPOSAL REQUIREMENTS and PROCESS

The CCPNR and CNRC contact is:

Christine Da Prat, Executive Director, CCPNR
Telephone: 902-388-4789
Email: ED@ccpnr.ca

Proposals are not to exceed five pages (single space and not including appendices) and as one PDF document. Proposals are to provide the following information:

1. contact name and information (telephone number, email address, organization address) of the proposed project lead;
2. brief and general description of expertise related to the project scope;
3. description of the proposed approach and methodology;
4. timeline for the project outlining key milestones;
5. estimate of the costs/budget;
6. identification and/or submission of similar research work and/or publications; and
7. three (3) references (name, position, organization name, email address and telephone number) including a synopsis of the assignment completed for that reference.

CCPNR shall not be obligated in any manner to any Respondent whatsoever. Respondents who choose to provide the information requested within this document do so at their own discretion and are solely responsible for their own expenses in preparing a response to this RFP.

This document should not be construed as an invitation for unsolicited proposals or an implied contract to purchase goods or services or any other procurement vehicle. CCPNR is not bound to purchase any goods or services as a result of a response to this RFP.

Confidentiality

The Respondent shall:

a. keep strictly confidential all information as a result of participating in this RFP; and
b. only use, copy or disclose such information as necessary for the purpose of submitting a response or upon written authorization of CCPNR.

Evaluation

The response to this RFP will be evaluated by the steering committee and according to the following criteria:

	Weight
Project Understanding	**10%**
Qualifications and experience	**35%**
Approach and methodology	**40%**
Level of effort, fee structure and budget	**10%**
References	**5%**
TOTAL	**100%**

The steering committee reserves the right to short-list respondents and request a teleconference meeting to respond to further questions about their proposal. Such a teleconference meeting with the respondent and steering committee is tentatively scheduled for *August 29, 2018*. Respondents will be advised one business week in advance.

Notification

Respondents will be informed of the results of the RFP by email no later than *September 4, 2018*. CCPNR will negotiate the terms of the contract with the successful Respondent on behalf of the CNRC. If negotiations do not result in a mutually agreeable contract, CNRC reserves the right to select another Respondent or withdraw the RFP altogether.

Deadline

Proposals are to be electronically submitted to Christine Da Prat at ed@ccpnr.ca by 12:00 ET August 17, 2018.

APPENDIX A

Draft report setting performance standards on three English language proficiency tests for licensure to practice as a nurse in Canada

Setting Performance Standards on Three English Language Proficiency Tests for Licensure to Practice as a Nurse in Canada

Report Prepared for the College of Nurses of Ontario

April, 2010

Professor Alister Cumming
Ontario Institute for Studies in Education
University of Toronto

Note: The information and views expressed in this report are those of its author and do not represent the policies of the College of Nurses of Ontario nor of the testing agencies involved except where so explicitly cited. Lindsay Brooks, a Ph.D. student in Second Language Education and instructor in the TESOL Certificate Program, Woodsworth College at the University of Toronto, assisted ably in preparing for and conducting the workshop.

EXECUTIVE SUMMARY

A standard-setting workshop was held in Toronto April 12 to 15, 2010 involving 17 representatives from regulatory agencies, educational programs, and professional practice for nursing across Canada, who agreed upon common standards for licensure of internationally educated nurses on three relevant English language proficiency tests: International English Language Testing System (IELTS), Canadian English Language Benchmark Assessment for Nurses (CELBAN), and Test of English for International Communication (TOEIC). For each test, the group reached a consensus on scores that represent the minimum competency in English to practice as a nurse in Canada. The workshop involved participants making analytic judgements—based on their informed, professional experience, then group consensus—to determine borderline abilities on sample tasks and items from each test. Participants recommended the following minimum scores on the writing, speaking, listening, and reading sections of each test:

	IELTS		**CELBAN***	**TOEIC***	
	Minimum band score	*SEM*	Minimum benchmark score	Minimum final score	*SEM*
Writing	7.5	0.5	7	200	N/A*
Speaking	7.5	0.5	8	190 to 200	N/A*
Listening	8	0.5	10	495	25
Reading	7	0.5	8	435	25
Total Score	7.5	0.5	N/A*	930	50*

*Note: The Standard Error of Measurement (*SEM*) is accounted for in the conversion from raw scores to final reported scores in all sections of the CELBAN and in the Writing and Speaking sections of the TOEIC. The *SEM* for TOEIC applies therefore only to scores from the Listening and Reading sections of that test (not the Writing or Speaking sections). CELBAN reports scores only for sections of the test rather than as total scores for the test overall.

The present report documents the procedures by which these standards were established, describing the relevant qualifications of the 17 participants and their observations about the standard-setting process, the scores established as minimum requirements on each test, as well as recommendations for uses of these scores in policies to licence internationally educated nurses to practice nursing in Canada.

Purpose and Scope

The present report documents the processes undertaken and results obtained during a series of workshops held April 12 to 15, 2010 in Toronto to establish minimally acceptable scores on three English language proficiency tests for licensure to practice as a nurse in Canada. The three tests are the International English Language Testing System (IELTS), Canadian English Language Benchmark Assessment for Nurses (CELBAN), and Test of English for International Communication (TOEIC). Logistics for the workshops were organized by the College of Nurses of Ontario. The workshops were facilitated by Alister Cumming, a professor at the University of Toronto who specializes in English language assessment, assisted by Lindsay Brooks, a Ph.D. student in language assessment and instructor at Woodsworth College, University of Toronto, as well as by Robert Edwards, a professor at the University of Sherbrooke, who has prepared a parallel report on the standards established for proficiency in French, on April 16, in reference to the *Test d'Evaluation de Français* (TEF). The first part of the report describes the procedures for standard-setting undertaken. The second part of the report describes the standards determined for each of the three tests. The final part of the report discusses the process and relevant policy issues.

Procedures for and Context of Standard Setting

Alister Cumming responded to a request in October 2009 from Diane Allen of the College of Nurses of Ontario to prepare, conduct, and then report on a series of standard-setting workshops on several English language tests for the nursing profession in Canada. Professor Cumming and Ms. Allen conferred on the composition of a panel suitable to establish these performance standards on a national basis, feasible dates to hold a series of workshops, and on an additional, appropriate facilitator for the French language test (TEF), Robert Edwards. Ms. Allen and colleagues arranged to invite to the workshops in

mid-April a selection of participants from relevant regulatory authorities, educational programs, and nursing support programs with interests in the certification of internationally educated nurses from across English-dominant Canada.

Professor Cumming organized the workshops to follow conventional procedures commonly used for setting minimally acceptable performance standards on tests for certification or credentialing purposes (Hambleton, 2001; Plake & Hambleton, 2001; Zieky & Perie, 2006) as well as procedures he had used in previous, similar projects (e.g., Cumming, 2009), all following principles recommended by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) in their *Standards for educational and psychological testing* (1999, pp. 53-54, 59-60). Specifically, the present workshops followed the "contrasting groups" and "borderline" methods (Cizek, 2001, p. 12; Zieky & Perie, 2006, pp. 15-17). These approaches to standard-setting involve analytic judgements by appropriately qualified panelists (i.e., in the present case, representatives from regulatory agencies, educational programs, and professional practice related to nursing across Canada) on benchmark samples of examinees' performances on tasks and items from each test (provided in this case by each of the three testing agencies) under the guidance of a facilitator knowledgeable about English language assessment (Professor Alister Cumming). For the present standard-setting activity, panelists determined, individually then discussed collectively to reach a consensual agreement, which performances on the sample test tasks or items represented their estimations of minimally acceptable levels of English language proficiency for licensure to practice as a nurse in Canada. As Hambleton (2001, pp. 92-93) observed, this method involves identifying "a single performance standard" or "cutoff score" that "separate[s] examinees into two performance categories…*certifiable* and *not certifiable*."

Hambleton (2001, pp. 93-94) describes this procedure as involving 11 steps, each of which were followed in the present case: (1) choosing an appropriate, representative panel of judges, (2) selecting an appropriate method for standard-setting and preparing orientation materials, (3) producing descriptions of the tests and performance categories, (4) orienting panelists to use the method, (5) having panelists rate the "performance of examinees at the borderlines of the performance categories" and compiling these ratings,

(6) conducting discussions among panel members and providing "feedback on interpanelist and intrapanelist consistency", (7) compiling ratings "a second time followed by more discussion, feedback, and so on", (8) compiling panelists' ratings to determine the performance standards, (9) presenting consequences data to the panelists, (10) revising if necessary and finalizing the performance standards and evaluating panelists' confidence in the resulting standards, and (11) compiling evidence and technical documentation (i.e., the present report).

Participants

The 17 participants in the standard-setting workshops represented relevant regulatory authorities, nursing education programs, nursing practitioners, and experts working with internationally educated nurses from 8 provinces and English-dominant regions across Canada. Their names, primary work responsibilities, and professional and/or regional affiliations are:

1. Diane Allen, Policy Analyst, College of Nurses of Ontario
2. Kathryn Allen, Nurse Educator, RN–Professional Development Centre, Nova Scotia
3. Angeles Amoyan-Bell, Community Nurse, Victorian Order of Nurses–Canada, Toronto York Branch
4. Sandra Berman, Vancouver Coastal Health Project Manager, College of Registered Nurses of British Columbia
5. Lois Berry, Associate Dean (Acting), North and North-Western Campus, Rural and Remote Engagement, College of Nursing, University of Saskatchewan
6. Jean Farrar, Deputy Registrar, College and Association of Registered Nurses of Alberta
7. Shelley Goodman, IEN Assessment Centre Coordinator, School of Nursing, Mount Royal University, Alberta
8. Odette Comeau Lavoie, Consultant on Regulatory Services, Nurses Association of New Brunswick

9. Ann Mann, Executive Director/Registrar, College of Licensed Practical Nurses of Nova Scotia
10. Kate Mercer, Nursing Instructor, Registered Nurses Professional Development Centre, Capital District Health Authority, Nova Scotia
11. Carole Reece, Nurse Consultant, IEN Assessment Centre, Saskatchewan Institute of Applied Science and Technology
12. Maria Theresa Refuerzo, Staff Nurse, CARE Centre for Internationally Educated Nurses, Toronto, Ontario
13. Khairrunnissa Rhemtulla, Manager, Learning and Development, Vancouver Coastal Health
14. Colleen Semmler, Registration Advisor, Internationally Educated Nurses, College of Registered Nurses of Manitoba
15. Della Shouse, Recruitment and Retention Officer (Nurses), Labrador-Grenfell Health, Labrador
16. Linda L. Stanger, Executive Director, College of Licensed Practical Nurses of Alberta
17. Karen Webber, Associate Director, Undergraduate Programs, School of Nursing, Memorial University, St. John's, Newfoundland

All participants, including the workshop facilitators, signed agreements declaring they would maintain confidentiality about the process and content of the workshops (except in reference to this final report, and only after it is officially accepted). Establishing confidentiality among participants helped to assure that participants could express their views freely during the workshops while also preserving the security of the test materials discussed, many of which involved samples from actual tests. Completed confidentiality agreements about the workshop processes and results are held at the College of Nurses of Ontario, and a second set of confidentiality agreements regarding the test materials, along with all other materials on which individual benchmark decisions were recorded during the workshops, are stored in a locked filing cabinet in Professor Cumming's office at the University of Toronto.

Tests Considered and Context of Standard-setting

The three English language proficiency tests considered for standard setting had been selected, prior to the April workshops, by a national group representing regulatory authorities for the nursing profession in Canada. The three tests are the International English Language Testing System (IELTS), Canadian English Language Benchmark Assessment for Nurses (CELBAN), and Test of English for International Communication (TOEIC). Diane Allen explained at the start of the workshops that the guiding purpose of the present standard-setting is to establish common standards across Canada for uses of these tests in licensing internationally educated nurses to practice in Canada. Ms. Allen also explained how the three tests had been selected in respect to several criteria: The three tests are comprehensive in the sense of each involving assessments of individuals' writing, speaking, reading, and listening abilities—providing separate scores for each of these four aspects of English proficiency. The tests are widely accessible to test-takers across Canada as well as internationally (for applicants overseas) and they are administered frequently. The tests each have established reputations for their psychometric characteristics and administration (i.e., validity, reliability, security, and score reporting) and have been used for some time, with satisfaction, by nursing regulatory authorities and related educational programs in Canada. The tests also represent a range of different cultural and work-related orientations for adults seeking certification of their English abilities: CELBAN was developed for and oriented to nurses in Canada, IELTS was developed for higher education and immigration in the United Kingdom and Australia (and more recently has been recognized for immigration applications by Citizenship and Immigration Canada), and TOEIC was developed for work situations in English in Asia as well as the United States and Canada. Alister Cumming, Robert Edwards, and Lindsay Brooks agreed with these descriptions of the three tests and their relevance to the present context.

Establishing a Common Basis for Setting English Proficiency Standards

At the beginning of the workshops, the facilitators encouraged participants to specify and agree upon the purpose and context of the standard setting. Participants agreed that their primary purpose was to set standards as scores on each of the three tests that represent minimally acceptable levels of English proficiency to be licensed to practice nursing in Canada. That is, the standards to be established emphasize a regulatory perspective—of certification for internationally educated and/or experienced nurses to work competently as nurses in English in Canada—rather than to function as a requirement for entry into a nursing, English, or other educational program. In making this distinction (between certification to practice and entry into an educational program), the participants also acknowledged the great variability in roles and situations in which practicing nurses may work—in situations of varying cultural and linguistic diversity, ranging from direct care, interactions with other health professionals and families, administration, as well as instructional roles. Participants also recognized that the three English proficiency tests assess individuals' language abilities generally rather than tests, such as the Canadian Registered Nurses Exam, that are designed to assess knowledge and skills explicitly for nursing (rather than English language). At the same time, participants asserted that it is important for the tests to report separate, reliable scores for each major aspect of English communication (i.e., speaking, writing, listening, and speaking) because each aspect is integral to effective practice as a nurse.

To establish a common basis for these decisions during the standard setting, the facilitators first prompted the participants to specify, based on their professional knowledge and experiences, the most important and consequential functions in which practicing nurses must communicate effectively in English in Canada. Participants wrote out their individual expectations and then discussed and agreed upon these collectively.

For writing, participants indicated tasks that involve: chart documentation, transcribing doctor's orders, preparing medication documentation, giving instructions to patients and families, care planning, note-taking in health assessment interviews and in emergency situations, filling out forms, referrals and reports for inter-professional communications, communicating with professional associations and following legal principles, and preparing shift reports—all in correspondence that may involve electronic

and handwritten formats, letters, short reports, and continuing development of these competencies.

For speaking, participants indicated that practicing nurses must have effective spoken communications: with clients, patients, families and health care teams, both therapeutically and professionally; for teaching and discharge planning; while receiving orders from physicians and others by phone, and during rapidly changing or emergency events; during interviews with employers, patients, professionals, and in history taking; and interacting in meetings. These speaking abilities primarily involve being able to articulate ideas so they are understood as intended.

For reading, participants cited reading and interpreting accurately: doctor's orders, referrals from other specialists, medical abbreviations, and patient-client records; research in looking up tests, procedures, policies, findings, and other evidence or information; remaining current with health literature, practice guidelines, and policies; understanding non-specialist information, such as directions, units of measure, using technologies, either electronically or in paper-based media; what needs to be done or actions taken with a clear understanding; and doing so with efficient speed and abilities in urgent, extreme situations as well as reading for depth and breadth.

For listening, participants cited the same types of tasks as for reading but also nurses' abilities to understand orally local, cultural references in context, including multicultural contexts, involving diverse accents, language varieties, and cultural norms; interpreting oral messages as well as comprehending non-verbal communications, requiring active listening, above and beyond basic language skills, in person as well as through telephone media, in a range of situations from emergencies to ordinary routines, and linked with ongoing speaking and conversational interactions. The participants emphasized the importance and consequences of all of these aspects of English communication but highlighted listening abilities as especially crucial in nursing practice.

The workshop facilitators then prompted the participants to specify further these aspects of English communication integral to nursing, as a basis of common reference for judging performance on the three English language proficiency tests, by writing out, then collectively agreeing upon, the characteristics of communication that are minimally acceptable to practice as a nurse in Canada in English. For writing, participants

indicated: legibility (especially, precise use of numbers and letters), accuracy and grammar, punctuation, spelling, consistency and flow of written discourse, appropriate vocabulary, and knowledge of document standards. For speaking, participants indicated: clear and understandable pronunciation, appropriate tone of voice, speed of speech, and succinctness; accuracy in uses of tenses, pronouns, and other grammatical features; abilities to use nuances, idioms, expressions, and terminology; adapting appropriate registers for contexts; knowledge and ability to use effective communication strategies (e.g., negotiating, asking for clarification, or paraphrasing); abilities to give directions and instructions clearly and appropriately for the audience; and adaptability to diverse situations. For reading, participants cited efficient speed, comprehension (e.g., identifying and extracting key concepts as well as major and minor points); and abilities to read and interpret a variety of texts of different registers and difficulty. For listening, participants cited similar expectations as for reading but also: abilities to understand phrases, idioms, colloquial expressions; comprehension (identifying key concepts and distinguishing major and minor points); abilities to listen to and interpret a variety of texts in different registers; displaying understanding after listening (e.g. by paraphrasing, repeating back instructions for confirmation); abilities to seek clarification while interacting with others; and abilities to cope with people in a variety of states of health (e.g., patients who have had a stroke, tracheotomy, or other impairments) and of diverse cultural and professional backgrounds.

Procedures for Standard-Setting

For the remainder of the standard-setting workshops, the facilitators introduced each test by describing its main characteristics and format then led participants to make judgments about minimally acceptable scores on the sections of the tests by listening to and/or reading and reflecting on samples of the tests and examinee performance on the tests provided by the test developers. The sequence of presentation during the workshops was IELTS, CELBAN, and then TOEIC. Approximately a full day was spent with each test, segmented into periods of about two hours for each of the four sections of each test

(each presented in the sequence of speaking, writing, listening, and then reading, except for the CELBAN which combined writing and listening tasks in the test).

For speaking, participants heard recorded speech samples of real examinees performing tasks from the tests at English proficiency levels below, at, and above the probable proficiency levels that might be considered as minimally acceptable cut scores on the tests used by the nursing profession. Each participant was asked to make an individual judgment about each speech sample, using a prepared judgment sheet to record this decision. Next the whole group of participants exchanged and discussed their reasons for their individual decisions, aiming under the guidance of the facilitators to reach a consensus on a mutually agreed upon cut score on that section of the test for internationally educated nurses in Canada. In some instances, a full consensus was reached by mutual agreement, and in other instances an average was taken from the range of scores determined by individuals. A similar process was followed for writing, except that participants read samples of written compositions or tasks that real examinees had performed on the tests, followed by individual decisions, group discussions, and efforts to reach a consensus.

The listening and reading sections of the tests do not produce samples of examinee performance, but rather they consist of items that involve comprehension of a pre-recorded oral interaction or of a written passage to which examinees respond to questions or other item types (e.g., multiple-choice, true-false, and other test formats) that assess their comprehension of specific ideas, language, or other aspects of the source material. To determine minimally acceptable scores on the listening and reading sections of the tests, participants were first asked to take these sections of the tests themselves (i.e., as examinees would). Then for each subsection of the test, participants were asked to make individual judgments as to whether they thought an internationally educated nurse should, or should not, be able to answer the item or complete the task correctly. Again, these judgments were documented on individual judgment sheets. Then moving through each subsection of each test, the workshop facilitators prompted the participants to reach a mutual consensus on each item or task then compiled participants' total expected scores, tallying the range from maximum to minimum as well as the average for all 17 participants.

In the final afternoon of the workshops (April 15), the facilitators presented a table of the results obtained for all three tests (as shown in the next section of this report), and engaged the participants in discussion of their scores. Specifically, participants were invited to consider and debate whether they wished to modify their individual or group judgments on the basis of having (a) now become familiar with all three tests, (b) seen the preliminary results of their standard-setting across the three tests, and (c) observed minimum, maximum, and average scores for the group of participants. For the most part, participants reconfirmed their initial judgments about cut scores appropriate for each test, but in a few instances minor modifications were made after discussion of this evidence. That is, a few individual participants agreed to modify their initial score judgments where they observed these to be distinctly different from the group of participants as a whole, particularly where average scores were close to a round number or crucial score point on the test. Participants observed that their minimum recommended scores for listening abilities were consistently higher, for each test, than were their comparable recommended scores for writing, speaking, or reading. In view of this trend, participants reaffirmed their decisions and their agreed-upon perceptions of the importance and consequences of listening abilities in English for nursing practice.

Standard Error of Measurement

Consideration, during these final discussions, was also given to the Standard Error of Measurement (SEM) for each test, a topic that had been briefly raised while considering minimum scores for each test. The SEM is the range within which scores on different administrations or versions of each test tend to vary, providing an estimation of the reliability of the scores on a test (which are never perfect on any test of human abilities). The workshop facilitators emphasized how the SEM needs to be accounted for when interpreting reports of scores from each test. For this reason, consideration of the SEM is a factor relevant to establishing the range of cut scores determined to represent minimal competency for each section of each test. The importance of considering the SEM in the present context is that when regulatory authorities interpret the minimum cut-point scores for individual test-takers, they should account for this fluctuation in scores

rather than treating the cut-point score as a fixed, absolute number. The SEM simply reflects an inherent limitation in the technology of test administration (i.e., measurements are never exactly precise, particularly for complex abilities like language proficiency). For this reason, most reputable testing agencies regularly analyze—and encourage people who use score reports for high-stakes decisions to make use of—the results across different administrations of their tests and publish information about the resulting SEM, following professional guidelines and expectations (e.g., Standard 5.10, AREA, APA & NCME, 1999, p. 65; Stoynoff & Chapelle, 2005, pp. 118-119).

The workshop facilitators presented information about the SEM for each test, based on reports from each testing agency. The facilitators also explained how these calculations differ for each test, depending on the nature of the tasks on the tests as well as the scales used to score examinees' results and to convert these raw scores to benchmark or proficiency levels that appear in final score reports. An explanation of the SEM is presented below in describing the recommended minimum scores for each of the three tests because the methods of calculation differ for the IELTS, CELBAN, and TOEIC. Such differences should be expected because each of these tests puts into operation the concept of English language proficiency differently, using different tasks, items, and scoring scales, which in turn produce different methods for estimating or accounting for the SEM.

For the sake of clarity, however, consider an initial example here, using the reading and listening sections of the TOEIC because these are the most longstanding of the tests considered at the workshops, and their statistical properties such as the SEM are well established (cf. Schmitt, 2005). The Technical Manual for the TOEIC states that the "standard error of measurement (SEM) for the Listening Comprehension section scale score is about 25 scale score points (1 SEM = 25 scale points) and is approximately the same for the Reading Comprehension section" (p. IV-5, Educational Testing Service, 1998). During the present standard-setting workshops, participants agreed upon a minimum required score of 435 for the reading section of the TOEIC. The SEM for this section of this test is 25. So scores that are either 25 below or 25 above the cut-point score of 435 must be considered equivalent on this test. That is, scores ranging from 410 to 460 on the reading section the TOEIC need to be interpreted as equivalent, given the

range of precision for this section of the test, and be considered to have fulfilled the minimum competency requirement. This SEM of 25 points is the range that the test developers estimate scores on any one administration of the test tend to vary from a probable "true score" that the test is capable of producing (across different administrations of the test).

Recommended Minimum Scores for the Three Tests

Table 1 displays the scores that the 17 participants in the standard-setting workshop recommended be established as the minimum required scores for the IELTS, CELBAN, and TOEIC for an internationally educated nurse to be certified to practice as a nurse in English in Canada. The subsections that follow describe the characteristics of each test and document issues unique to the standard-setting processes and recommended scores for that test.

Table 1. Recommended Minimum Scores on Three English Proficiency Tests for Nursing Practice in Canada

	IELTS		**CELBAN***	**TOEIC***	
	Minimum band score	*SEM*	Minimum benchmark score	Minimum final score	*SEM*
Writing	7.5	0.5	7	200	N/A*
Speaking	7.5	0.5	8	190 to 200	N/A*
Listening	8	0.5	10	495	25
Reading	7	0.5	8	435	25
Total Score	7.5	0.5	N/A*	930	50*

*Note: The Standard Error of Measurement (SEM) is accounted for in the conversion from raw scores to final reported scores in all sections of the CELBAN and in the Writing and Speaking sections of the TOEIC. The SEM for TOEIC applies therefore only to scores from the Listening and Reading sections of that test (not the Writing or Speaking sections). CELBAN reports scores only for sections of the test rather than as total scores for the test overall.

IELTS

Participants in the workshop made clear that they wished to set standards in reference to the Academic, rather than General Training, Modules of the IELTS, given

the language demands and expectations for academic literacy expected of nurses in Canada. University of Cambridge ESOL Examinations (2009) produced a relatively high quality set of materials in DVD format specifically for standard-setting purposes around the world, such as the present workshops, so the facilitators decided to begin the workshop with these materials. The materials give samples of tasks, examinees' performance, and evaluators' comments at each score level on each section of the test. For the writing and speaking sections of the IELTS, examinees' performance on several holistic tasks is rated directly in increments of half score points (i.e., 0.5) on 9 bands. The listening and reading sections of the IELTS, in contrast, consist of multiple items, so a raw score is initially calculated as the sum of correct items for these sections and then converted to a band score, so as to make a scale equivalent to the band scores on the writing and sections of the test, except that scores for the reading and listening sections of the IELTS are reported on scales in whole numbers (e.g., 6, 7, 8, or 9, rather than half-band points). The total score for the IELTS is an average of the section scores on the test.

Table 2. Section and Total Band Scores Recommended for IELTS, *SEM*, and Descriptive Statistics for Participants' Initial Individual Judgments (*N* = 17)

Test	Section	Min.	Max.	*M*	*SD*	Band/Score	*SEM*
IELTS	Writing	7	8	7.4	0.37	7.5	.5
	Speaking	7	8	7.3	0.31	7.5	.5
	Listening	32	40	36.0	2.35	8	.5
	Reading	24	34	28.2	3.07	7	.5
	Total					7.5	.5

As Table 2 shows, participants in the standard-setting workshop decided that minimum scores on the IELTS for licensure to practice as a nurse in Canada should be: 7.5 for Writing, 7.5 for Speaking, 8 for Listening, 7 for Reading, and 7.5 for the IELTS overall. Interpretation of these minimum scores on individual score reports should take into account the published *SEM* for each section of the IELTS, which is .5 on overall band scores (University of Cambridge ESOL Examinations, 2006). So, for example, a score of 7 on the Writing section of the IELTS is within the *SEM* of .5 for the minimum required score of 7.5, and so should be interpreted to be within that range (given the

estimated reliability of the test). In contrast, scores of either 6.0 or 6.5 would not be within that range and so should not be considered to fulfill the required minimal competency score.

As also shown in Table 2, participants were in relatively close agreement in their individual judgments about minimum scores on each section of the IELTS. Participants were almost evenly split between recommended scores of either 7 or 8 for the writing and speaking sections of the IELTS, so they agreed upon a required score of 7.5 (and those tending to prefer a score of 7 acknowledged this would, if effect, be the required score in view of the *SEM*). There was little variation (as shown by the standard deviations in Table 2) among individuals between the recommended scores on the listening and reading sections of the IELTS, so participants readily agreed to accept the average of the recommended scores. Participants recognized that their expected scores for the listening section are higher, and their expected scores for the reading section are lower, than for other sections of the IELTS, and they affirmed that is their intention in view of the performance required for these sections of this test and their collective understanding of the abilities required for nursing practice in Canada.

Overall, these recommended scores are well within the range of IELTS' own recommendations of scores between 7 and 9 for academic programs that are either "linguistically demanding (e.g., medicine, law, linguistics, journalism)" or "linguistically less demanding (e.g., agriculture, pure mathematics, technology, IT, and telecommunications" (University of Cambridge ESOL Examinations, 2010, p. 10). However, the present recommended scores are slightly higher than the IELTS scores between 6 and 7.5 that O'Neil, Buckendahl, Plake, and Taylor (2007, Table 8, p. 311) cited as established minimal requirements for various associations of health professionals in the U.S., Canada, Australia, and New Zealand.

CELBAN

The CELBAN was developed uniquely to assess the English language abilities of internationally educated nurses in Canada, following an extensive needs analysis about a decade ago (Epp, Stawychny, Bonham, & Cumming, 2002) then test development and

implementation in subsequent years at Red River College in Winnipeg and through the Centre for Canadian Language Benchmarks in Ottawa. The initial needs analysis had recommended minimum expectations for English proficiency of nursing professionals in respect to descriptions in the *Canadian Language Benchmarks* (2000) as 7 for writing, 8 for speaking, 9 for listening, and 8 for reading. These recommendations were made, however, before the CELBAN was actually developed or implemented, so it seemed to participants in the present workshop worth establishing new minimum competency levels now in view of their judgments about samples of the CELBAN test materials. Participants asserted this should be done in respect to the official version of the CELBAN, considering its purposes of certifying nurses for licensure, rather than the institutional version of the CELBAN, which is designed for purposes of determining placement or progress within educational programs for internationally educated nurses.

Test developers at Red River College provided the workshop facilitators with sample copies of CELBAN test materials and examinees' performance on speaking and writing tasks in print and DVD formats as well as the *CELBAN administrators and invigilators guide* (2009*)* for the limited purpose of the present standard setting. The CELBAN was developed to represent realistically communication tasks typical of nursing practice in Canada, making it a specific-purpose language test (Douglas, 2000). In this fundamental respect, the CELBAN differs from the IELTS or TOEIC, which were developed to assess adults' English abilities generally across a range of situations in work-place and academic communications in various fields and situations. The speaking and writing tasks for the CELBAN are scored with detailed, complex scoring schemes, which are then converted to final scores that represent benchmark levels on the *Canadian Language Benchmarks* (2000). Because of the complexity of this scoring, which requires specialized training, participants in the present workshop were simply given samples of examinees' performance (previously scored to represent relevant benchmark levels) in speaking and writing and were then asked to judge holistically which of these they considered minimal scores for nursing practice, as was done for comparable sections of the IELTS and TOEIC. As with the IELTS and TOEIC, participants themselves did the reading and listening sections of the CELBAN and made estimates for each item on these

sections of the test as to whether internationally educated nurses should, or should not, be expected to perform successfully on the items.

To the best of our understanding, the conversion formula (that converts the detailed raw scores on the CELBAN into the final scores expressed in terms of Canadian Language Benchmark levels) accounts adequately for the *SEM* on this test (and this process is akin to the conversion formula used for TOEIC speaking and writing tasks, see below). Developers of the CELBAN at Red River College reported to us, upon request for the present workshops, an *SEM* of about 3 raw score points across different sections and versions of the test. But this small number of 3 raw score points is subsequently rounded out when the raw scores are converted to the smaller number of 12 benchmark levels from the Canadian Language Benchmarks. The upshot is that users of scores from the CELBAN should not interpret individual scores from the CELBAN in relation to an *SEM* and, indeed, there is not an empirical basis for doing so when the final, reported scores are expressed as proficiency levels from the Canadian Language Benchmarks.

Table 3. Section and Total Benchmark Scores Recommended for CELBAN and Descriptive Statistics for Participants' Initial Individual Judgments (*N* = 17)

Test	Section	Min.	Max.	*M*	*SD*	Benchmark Score
CELBAN	Writing	7	8	7.4	0.5	7
	Speaking	8	8	8	0	8
	Listening	77	82	80.9	1.59	10
	Reading	49	86	70.5	9.03	8

Table 3 shows that participants in the standard-setting workshop decided that minimum scores on the CELBAN for licensure to practice as a nurse in Canada should be: 7 for Writing, 8 for Speaking, 10 for Listening, and 8 for Reading. The CELBAN only reports section scores rather than a total score for the test. There was perfect unanimity among participants for the recommended score of 8 for speaking abilities. Participants were split between recommended scores of 7 or 8 for the writing section of the CELBAN, and they agreed to opt for the more lenient score of 7. There was little variation among participants in the recommended score of benchmark level 10 for listening on the CELBAN. Participants recognized that this score is higher than the score

of 9 recommended in Epp, Stawychny, Bonham, and Cumming (2002), and all but a few participants affirmed that the score of 10 seemed appropriate in their judgments. There was considerable variation, however, among individual judgments (standard deviation of 9 in Table 3, ranging from a minimum of 49 to maximum of 86) in raw scores on the reading section of the CELBAN, but after discussion participants agreed that the resulting recommended benchmark score of 8 (after conversion from raw scores) was appropriate and represented their collective judgments. In sum, the recommended minimum scores for the CELBAN here remain the same as those recommended initially on the basis of surveys, interviews, and observations by Epp, Stawychny, Bonham, and Cumming (2002) except for scores for listening comprehension on the CELBAN, which the present participants asserted should be higher (at benchmark level 10) than the benchmark level 9 previously recommended.

TOEIC

Procedures for the TOEIC followed materials produced by Educational Testing Service for standard-setting on the relatively new speaking and writing sections of the TOEIC (Educational Testing Service, 2007), the Technical Manual for the relatively longstanding reading and listening sections of the TOEIC (Educational Testing Service, 1998), the handbook for orienting examinees to prepare to take the TOEIC (Educational Testing Service, 2006), and a sample reading and listening test in print and DVD formats provided, upon request from the workshop facilitators, for the present purposes by Educational Testing Service.

As described above, the TOEIC Technical Manual cites the *SEM* for the reading and listening sections of this test as 25, but observing that this *SEM* varies slightly according to the points on the scoring scale (i.e., the TOEIC is most precise and reliable in the mid-range of scores, as with most standardized tests). For the speaking and writing sections of the test, first implemented in 2007, however, scores are made first as holistic judgements on scales of 4 or 5, then these raw scores for each task are combined and then converted to a larger scale that represents normative performance by examinees on the test. This combination of raw scores then conversion to a final reported score is akin to

the process described above for the CELBAN (though the process of converting between raw scores and scales goes in reverse for the writing and speaking sections of the TOEIC from small initial raw scores to larger, final scaled scores). The issue to emphasize is that this conversion process rounds out, and thereby accounts implicitly for, the *SEM* for the scores on the speaking and writing sections of the TOEIC. So users of scores from the TOEIC should not expect to interpret scores for the speaking or writing sections of the TOEIC in respect to an *SEM* (because the score conversion process has already done that). But users of scores from the TOEIC should expect to interpret scores from the reading and listening sections of the TOEIC in respect to an *SEM* of 25 for each of these sections. Conceptually, the newer speaking and writing sections of the TOEIC are based on criterion-referenced models of assessment, whereas the older reading and listening sections of the TOEIC are based on norm-referenced models of assessment (cf. Brown & Hudson, 2002).

Table 4. Section and Total Scores Recommended for TOEIC and Descriptive Statistics for Participants' Initial Individual Judgments (*N* = 17)

	Section	**Min.**	**Max.**	***M***	***SD***	**Proficiency Level**	**Score**	***SEM***
TOEIC	Writing	4	5	4.71	.47	9	200	X
	Speaking	5	5	5	0	8	190 to 200	X
	Listening	80	100	96.4	5.67	X	495	25
	Reading	64	98	89.7	7.99	X	435	25
	Total						930	50

Table 4 shows participants in the standard-setting workshop decided that minimum scores on the TOEIC for licensure to practice as a nurse in Canada should be: 200 for Writing, from 190 to 200 for Speaking, 495 for Listening, 435 for Reading, and 930 for the TOEIC overall. The total score for the TOEIC is the sum of the scores for the reading and listening sections combined; scores from the writing and speaking sections are not included in this total. The range from 190 to 200 for the speaking section of the test is the way in which the TOEIC developers represent the score point decided upon by participants in the present workshop; presumably it is represented as a range rather than a fixed score point to reflect the relative imprecision in this aspect of the test.

Participants in the workshops were able to reach agreement readily on recommended scores for the listening and reading sections of the TOEIC, and they expressed confidence in these components of the test, despite variation in their scores collectively for the reading section (as also appeared for the reading section of the CELBAN). Table 4, however, represents the figures about the writing and speaking sections of the TOEIC in shading because participants in the workshops were not fully convinced that they could set standards for these two aspects of the TOEIC. This was a singular point of controversy and uncertainty during the four days of standard-setting.

Three related problems emerged and intersected. One problem was that the samples of speech and writing in the *Propell Workshop* materials (Educational Testing Service, 2007) were too few in number and range and, for the speaking task at the most crucial score point, too poor in audio quality to permit participants to make firm judgments about people's minimal competency in English. Moreover, only one speech or writing sample was provided for each score point, and participants felt this was an inadequate sample on which informed judgments could be made. A second, related problem was that some participants in the workshop did not think that the speech and writing samples in these materials corresponded accurately to the descriptive criteria published for the test. In discussing this matter, concern was expressed that the speech and writing samples may be relevant to contexts where English functions as an international medium of communication rather than the dominant language in society, reflecting the development and widespread use of the TOEIC in Japan and Korea (cf. Schmitt, 2005), rather than norms or expectations for professional purposes in English-dominant Canada. The third problem was a concern that the highest levels of speaking and writing assessed on the TOEIC may actually be lower than the level of English proficiency expected for nursing practice in Canada, particularly when considered in comparison to similar materials for the IELTS and CELBAN. Fundamentally, is the TOEIC aimed at an English proficiency level below that expected for nursing practice in Canada? In response to this concern, the workshop facilitators played audio samples of examinees' performance on speaking tasks (similar to those on the TOEIC) from the standard setting materials (Educational Testing Service, 2004) for the Test of English as a Foreign Language (TOEFL), a test also produced by Educational Testing Service but for

university admissions in North America and pitched at higher levels of English proficiency than is the TOEIC. However, given the aims of the present workshop to set standards for the TOEIC, there was neither ample time nor a policy basis to consider whether the TOEFL might be a more suitable test than the TOEIC, on a par with English language proficiency norms in the IELTS and CELBAN, for the purposes of licensing nurses in Canada.

Participants' Observations on the Process

At the conclusion of the four days of workshops, panelists were asked to write out their impressions of the standard-setting process individually and anonymously, evaluating whether they found the process effective, were confident in the decisions reached, and could offer suggestions for future sessions. The 16 responses (from 17 participants) seemed frank and sincere.

All participants stated they found the standard-setting process effective, citing: the time and opportunities provided to discuss test materials, the thorough processes to achieve group consensus, the cross-section of nursing professionals represented from across Canada, participants' concerns for maintaining high standards of professionalism for nurses in Canada, the sequence and flow of activities, and the expertise and style of the facilitators. The few comments offering criticisms stated: they would like to have reviewed more samples of examinees' performances on each test, particularly for speaking, received more precise directions on the goals to be achieved during the workshops, had fuller orientation to the individual backgrounds and experiences of participants, taken more time to review, evaluate, and reach decisions to alter individual judgments on minimum scores, and set standards for entry into various types of educational programs as well as for licensure to practice nursing.

Similarly, all participants indicated they were confident in the decisions reached during the standard-setting workshops. Reasons cited for this confidence were: the thoroughness and thoughtfulness of the standard-setting process, the active, respectful, and collaborative involvement of all participants, the quality of questions posed and issues discussed, ample opportunities to express individual opinions, correspondences

made between professional experiences and analyses of test content, the inclusion of several internationally educated nurses among the participants, the specificity of the test materials presented, and the facilitators' guidance in and management of the process as well as summaries of discussions. Two people expressed concerns over the final score of 10 on the listening section of the CELBAN, and one of these people also expressed uncertainty over the sample materials provided for the TOEIC, though the person was pleased that time was taken to discuss concerns about the TOEIC thoroughly. Two people indicated that more internationally educated nurses could have been represented among the participants. One person was perplexed that issues about Standard Error of Measurement were introduced early in the process rather than in the final summary session. One person suggested that more nation-wide consultations of this kind should be organized to maintain a high level of professionalism among nurses in Canada.

Concluding Remarks

The present standard-setting workshops succeeded in establishing a consensus among the 17 participants on minimum scores for the IELTS, CELBAN, and TOEIC to certify that internationally educated nurses have sufficient proficiency in English to practice in Canada. These recommendations now need to be taken forward to establish common policies among relevant regulatory authorities across Canada. The participants agreed that standards for all aspects of English abilities need to be high for practicing nurses, and particularly so for listening comprehension. The participants seemed confident about the standards set and the processes undertaken to arrive at them. Some uncertainty remains about the suitability of the TOEIC for nursing licensure in Canada, though these concerns may reflect more on the quality of the standard-setting materials available for the new speaking and writing components of this test rather than the test's established components to assess listening and reading abilities in English. The standards set in the present workshops will be worth revisiting in another decade or so, as each of the tests may change as may expectations for language abilities related to professional practice in nursing. Participants in the workshops commented frequently, for example,

that they have observed the demands of literacy, communications, and technical skills required for nursing practice to have increased distinctly over the past decade.

References

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Brown, J.D. & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge: Cambridge University Press.

Centre for Canadian Language Benchmarks. (2000). *Canadian language benchmarks 2000: English as a second language—for adults*. Ottawa: Centre for Canadian Language Benchmarks.

Centre for Canadian Language Benchmarks and Red River College. (2009). *CELBAN administrators and invigilators guide (revised 2009).* Ottawa: Centre for Canadian Language Benchmarks.

Cizek, G. (Ed.) (2001). *Setting performance standards: Concepts, methods, and perspectives.* Mahwah, NJ: Erlbaum.

Cumming, A. (October, 2009). *Setting performance standards on four English language proficiency tests for licensure as a pharmacy technician in Canada.* Report submitted to the Ontario College of Pharmacy, Toronto, and the National Association of Pharmacy Regulatory Authorities, Ottawa.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge, UK: Cambridge University Press.

Educational Testing Service. (1998). *TOEIC technical manual*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2004). *Standard setting materials for the Internet-based TOEFL test*. DVD. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2006). *TOEIC examinee handbook*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2007). *Propell workshop for the TOEIC speaking and writing tests*. Princeton, NJ: Educational Testing Service.

Epp, L., Stawychny, M., Bonham, A., & Cumming, A. (2002). *Benchmarking the English language demands of the nursing profession across Canada*. Ottawa: Centre for Canadian Language Benchmarks and Red River College.

Hambleton, R.K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 89-116). Mahwah, NJ: Erlbaum.

O'Neil, T., Buckendahl, C., Plake, B., & Taylor, L. (2007). Recommending a nursing -specific passing standard for the IELTS examination. *Language Assessment Quarterly, 4*, 4, 295-317.

Plake, B., & Hambleton, R.K. (2001). The analytic judgment method for setting standards on complex performance assessments. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 283-312). Mahwah, NJ: Erlbaum.

Schmitt, D. (2005). Test of English for International Communication (TOEIC). In S. Stoynoff & C. Chapelle (Eds.) *ESOL tests and testing: A resource for teachers and administrators* (pp. 100-102). Alexandra, VA: TESOL.

Stoynoff, S. & Chapelle, C. (2005). *ESOL tests and testing: A resource for teachers and administrators*. Alexandra, VA: TESOL.

University of Cambridge ESOL Examinations. (February, 2006). IELTS test performance data 2004. *Research Notes, 23*, 13-15.

University of Cambridge ESOL Examinations. (2009). *IELTS scores explained (version updated 2009)*. DVD. Cambridge, UK: University of Cambridge ESOL Examinations.

University of Cambridge ESOL Examinations. (2010). *IELTS guide for educational institutions, government, professional bodies, and commercial organizations*. Cambridge, UK: University of Cambridge ESOL Examinations. Retrieved, April 24, 2010 from

http://www.ielts.org/PDF/IELTS%20Guide%20for%20Stakeholders%20March%202009.pdf

Zieky, M., & Perie, M. (2006). *A primer on setting cut scores on tests of educational achievement.* Princeton, NJ: Educational Testing Service.