NCSBN
National Council of State Boards of Nursing

111 E. Wacker Drive, Suite 2900
Chicago, IL 60601-4277
312.525.3600
www.ncsbn.org

Recommending a Nursing-specific Passing Standard for the IELTS Examination

Thomas R. O'Neill

National Council of State Boards of Nursing

Chad W. Buckendahl and Barbara S. Plake

Buros Center for Testing

University of Nebraska, Lincoln

Lynda Taylor

University of Cambridge ESOL Examinations, UK

Abstract

Licensure testing programs in the United States (e.g., nursing) face an increasing challenge of measuring the competency of internationally trained candidates, both in relation to their clinical competence and their English language competence. To assist with the latter, professional licensing bodies often adopt well established and widely available international English language proficiency measures.

In this context, the National Council of State Boards of Nursing (NCSBN) sought to develop a nursing-specific passing standard on the International English Language Testing System (IELTS) that US jurisdictions could consider in their licensure decisions for internationally trained candidates. Findings from a standard setting exercise were considered by NCSBN's Examination Committee in conjunction with other relevant information to produce a legally defensible passing standard on the test.

This paper reports in detail on the standard setting exercise conducted as part of this policy making process; it describes the techniques adopted, the procedures followed and the outcomes obtained. The study is contextualized within the current literature on standard setting. The latter part of the paper describes the nature of the policy-making process to which the study contributed and discusses some of the implications of including a language literacy test as part of a licensure testing program.

Recommending a Nursing-specific Passing Standard for the IELTS Examination

Background

Licensure testing programs in the United States (e.g., nursing) face an increasing challenge of measuring the competency of internationally trained candidates. These programs typically conduct practice analyses to define the job-related knowledge, skills, abilities, and judgments necessary for safe, entry-level practice. Because a large majority of candidates graduate from domestic training programs, there is an assumed level of English language literacy inherent in these programs. However, as new avenues emerge for accreditation of international training programs, this assumption may not extend to all candidates.

In the years 2001-2006, the number of first-time, NCLEX[1]-RN® candidates who were internationally educated has been increasing (Table 1). These clearly increasing numbers of nurse candidates who were educated outside the U.S. are even higher when repeat test-takers are included. For many internationally educated candidates, English is not their primary language. For nurse licensing boards that are responsible with regulating the practice of nursing in a manner that protects the public, this provides an additional challenge. Not only do boards of nursing need information regarding the clinical competence of the people applying for a nursing license, but each board also needs to know if the candidate has sufficient job-related English language skills to effectively employ their clinical abilities in the workplace. To assist with the latter, professional licensing bodies often make use of established and widely available international English language proficiency measures.

Given these trends and policy implications for language testing, NCSBN[2] has set out to assist its member boards with these issues. NCSBN is attempting to identify passing standards that reflect the minimum degree of English proficiency necessary to function safely and effectively as an entry-level nurse on commonly used English language proficiency tests. NCSBN previously conducted a standard setting study (O'Neill, 2004; O'Neill, Tannenbaum, & Tiffen, 2005; O'Neill, Marks, & Wendt, 2005) for the Test of English as a Foreign Language (TOEFL) (Educational Testing Service, 2003) as one of these commonly used tests. Continuing with this initiative, NCSBN has also considered additional tests to offer greater access to candidates who sought to practice in the U.S. Subsequently, NCSBN identified the IELTS examination[3] as another commonly used test on which to develop a recommended passing standard.

There are several factors that should be considered prior to selecting a test to measure language literacy. Prior to their adoption of the IELTS, NCSBN evaluated: (a) the nature of the test content and format (e.g. all four skill domains, the communicative and task-based approach, inclusion of a face-to-face, interactive speaking test), (b) the quality and reliability in test production and administration procedures, (c) the frequency and widespread test availability through a global network, (d) the fast turnaround of score reporting, (e) the availability of a wide variety of information and support materials for test users and test-takers, and (f) the extensive use and recognition of IELTS scores by other health professional regulatory and licensing bodies worldwide (e.g., UK General Medical Council, Nursing Council of Australia, American Association of Veterinary State Boards). The original NCSBN decision to adopt IELTS is not the focus of this paper so these aspects will not be discussed further here. It is worth noting, however, that the widespread use and recognition of IELTS scores by health professional regulatory and

---

[1] The NCLEX® Examination is the nurse licensing examination that is used in all 50 U.S. states as well as in American Samoa, the Virgin Islands, Guam, and the Northern Marianna Islands. There are two separate examinations: the RN for registered nurses and the PN for licensed practical nurses and vocational nurses.

[2] The National Council of State Boards of Nursing, Inc. (NCSBN) is a not-for-profit organization that is composed of 59 jurisdictional boards of nursing in the United States and US territories whose mission is to provide leadership to advance regulatory excellence for public protection.

[3] International English Language Testing System (British Council, IELTS IDP Australia, & University of Cambridge ESOL Examinations, 2003)

licensing bodies worldwide reflects the growing prevalence of English language testing that is occurring in addition to the content-specific elements of a given profession. Although in some contexts occupation-specific language proficiency tests for health professionals have been developed, e.g., the Occupational English Test (OET)[4] in Australia and the Canadian English Language Benchmark Assessment for Nurses (CELBAN)[5] in Canada, factors such as frequency of test administrations and distribution of test centers worldwide mean that less occupation-specific, international academic English language tests such as IELTS and TOEFL serve an important function in this regard.

After NCSBN's adoption of IELTS, the next step in the process was to establish the passing score(s) that candidates would need to demonstrate on the test. NCSBN convened a standard setting panel in Chicago on October 30-31, 2004. The purpose of this study was to gather information from panelists about the minimum passing score that nursing candidates would need to achieve on the IELTS. The results of the study were the panel's recommendations about the level of English language proficiency, as measured by IELTS, the panelists believed necessary for entry-level nurses to possess to be able to perform important nursing responsibilities safely and effectively. It is important to note that the standard is intended to reflect the minimum level of English proficiency necessary for safe and effective entry-level practice, not the level of proficiency necessary for nurse-candidates to take the NCLEX® examination.

Ultimately, the minimum passing score (MPS) is a purpose-specific policy decision that emerges from a performance level description of the expectations of target candidates. For example, the defined performance level for a news reporter or a trial lawyer may yield a MPS that is quite different than the MPS we might expect for a mathematician or computer programmer. This is a function of the role the English literacy construct plays in the job-related skills of the respective profession. Because the IELTS examination produces a band (scale) score, not a pass-fail decision, NCSBN set out to establish a recommended MPS specific to entry-level nursing. There are inherent advantages in making available a legally defensible passing standard to boards of nursing. First, rather than have the board of nursing in each US jurisdiction repeat the same study, it seemed more efficient to allocate substantial resources to the study and perform it just once. Second, the benefit to internationally educated candidates is that the examination results would be portable across all jurisdictions that use the standard.

The Process of Standard Setting

Psychometricians have developed a variety of approaches for identifying what a particular group of content experts consider to be the performance threshold between *minimally competent* and *incompetent*; several authors (Cizek; 2001; Kaftandjieva, 2004; Hambleton & Pitoniak, 2006; Cizek & Bunch, 2007) have provided detailed and useful overviews of the range of approaches and procedures that can be adopted. In Cizek (2001), Zieky comments that the act of determining a cut-score to reflect the performance threshold of interest "remains very much a matter of subjective judgment" (p. 21). The setting of standards is ultimately a policy decision, rather than any absolute or objective truth. The following example helps to illustrate the policy nature of establishing standards.

---

[4]The Occupational English Test (OET) is a language proficiency test for overseas qualified medical and health professionals whose first language is not English. Introduced in the early 1990s, it assesses English language competency as it is used in medical and health professions in Australia. The test is administered by the Centre for Adult Education (CAE) with support from the Language Testing Research Centre at the University of Melbourne. For more information: http://www.cae.edu.au/OET/

[5] The Canadian English Language Benchmark Assessment for Nurses (CELBAN) is an English language assessment recently developed by the Centre for Canadian Language Benchmarks (CCLB) specifically for internationally-educated nurses. The test is administered by the Canadian English Language Assessment Services Centre (CELAS). For more information: http://www.celban.org/

> "The ruler has been around for a long time and is generally regarded as a stable instrument for measuring distance. However when a child goes to an amusement park and asks why one must be a certain height to ride a particular ride, the explanation about the ruler's stability seems quite irrelevant. Why not an inch lower? Or higher? Of course, there is a safety-based rationale that considers acceptable risks behind the rule, but how safe a ride should be and what constitutes an acceptable risk are really personal judgments made by a person or a group of people" (O'Neill, Marks & Reynolds, 2005, p. 131).

Conceptually, a passing standard is a function of informed professional judgment that relies on the panelists' content expertise and their experience with the abilities of the target examinee population. There are no passing standards that are empirically correct. A passing score reflects the values of those professionals who participate in its definition and adoption, and different professionals may hold different sets of values. Its determination may be informed by empirical data or other information, but ultimately, the passing standard established by a policy-making body is a judgment-based decision. Other authors have also noted this issue with regard to cut-score decision (Kane, 1998: 137; Zieky, 2001).

The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) recommend that the rationale and procedures used to set a standard be clearly documented. This documentation should include a description of the standard setting procedure, the panelist selection process and the qualifications of panelists selected, as well as a description of the training provided. This is important because if the standard is challenged, clear documentation can provide evidence that the standard was not set in an arbitrary and capricious manner. Cizek (1993) likens standard setting to 'due process' in law and Camilli et al (2001) suggest the notion of *psychometric due process*, highlighting the importance of 'the proper following of a prescribed system of rules and procedures' (p. 450), so that these are transparent to all and open to scrutiny. Similarly, if the organization wants to have a standard adopted by others, the documentation of the rigorous procedures used to derive that standard helps to build confidence in that standard.

Method

IELTS examination

IELTS is an examination designed to assess English language ability in examinees for whom English is not their native language. The English that IELTS represents in the test is not restricted to a North American English model, but reflects standard English used in the international target destinations of most candidates, i.e., Great Britain, Ireland, New Zealand, and Australia, as well as North America. IELTS covers four different language skill domains, called modules (Listening, Reading, Writing, and Speaking) and reports a degree of proficiency in each skill using "band scores" that range from 0 to 9. These band scores are scaled by transforming the raw score performance separately on each module. Scores for the four skills are combined to produce an overall band score ranging from 0 to 9. The meaning of the overall band scores, as defined by IELTS, is provided in Table 2. However, it is important to note that these band score descriptions may or may not correspond to an agency's performance level descriptors that define the knowledge, skills, and abilities required for a specific target candidate population. Again, the definition of performance is unique to the intended use of the scores and the target population of examinees.

There are two different versions (Academic and General Training) of the IELTS that specifically apply to two of the modules. The Academic version is typically taken by candidates who are interested in pursuing higher education paths; the General Training version is more consistent with candidates who would need to use English in daily, functional activities, most often in non-academic, vocational training contexts. Given the type of education required for nurses (i.e. some level of higher education beyond high school), the Academic version was selected as being more congruent with the anticipated English usage demands of nursing

candidates. The Listening and Speaking modules are identical across the Academic and General Training versions of the test; however the Reading and Writing modules are slightly different given the different demands of candidates taking each version. Next, we will provide brief descriptions of the four modules that comprise IELTS in the order of their administration.

Listening module. This module has four sections, contains 40 questions, and takes approximately 30 minutes to complete. In this module, a recorded sample of speech is played for candidates; candidates read the questions and mark their answers. The first two sections of the module are based upon a dialogue and then a monologue related to social needs. The last two sections include a conversation with up to four people and then a monologue, both of which are related to educational contexts. The format of these questions can include multiple choice, short answers, sentence completion, chart/table completion, diagram labeling, classification, and matching.

Academic Reading module. This module consists of three passages, 40 questions, and takes 60 minutes to complete. The passages are taken from magazines, journals, books, and newspapers. All passages are written for a non-specialist audience and are considered to be of general interest. The three passages range from 2,000 to 2,750 words in all and may include graphic illustrations such as charts and diagrams. The format of these questions can include multiple choice, short answers, sentence completion, chart/table completion, select general theme (from a list), identify author's views (yes, no, or not given), identify information in the text (yes, no, or not given), classification, and matching.

Academic Writing module. This module contains two writing tasks and takes 60 minutes to complete. In the first task, the candidate is asked to write a minimum of 150 words to describe or explain a diagram or table. In the second task, the candidate is presented with a point of view or problem and asked to write a minimum of 250 words to support or refute the point of view or present a solution to the problem. Scoring rubrics have been developed to score the responses. The second task is weighted more heavily than the first task in the scoring for this module.

Speaking module. This module has three sections and takes 11-14 minutes to complete. It is a face-to-face interview with a trained examiner. The first section, which lasts four to five minutes, consists of the candidate introducing him/herself and then answering a few short questions. The second section consists of the candidate speaking for one to two minutes on a topic selected by the examiner; the candidate is given one minute of preparation time before their long turn at the end of which the examiner asks one or two follow up questions. The third section lasts for four to five minutes. In this section the examiner engages the candidate in a conversation about a topic linked to the Section 2 theme but more abstract in nature. Scoring rubrics have been developed to score the candidate's responses.

Procedures

Initially, the panel was led through a discussion about the English proficiency characteristics of the target candidate in the context of safe and effective entry-level nursing. This was supplemented with a discussion of the activities that had been identified by the *2002 RN Practice Analysis: Linking the NCLEX-PN Examination to Practice* (Smith & Crawford, 2003a) and the *2003 LPN/VN Practice Analysis: Linking the NCLEX-RN Examination to Practice* (Smith & Crawford, 2003b) as being within the scope of entry-level practice. After identifying critical nursing activities in which communication plays an important role, the panel was provided with training regarding their role in the standard setting exercise. Table 3 summarizes the results of the panel's discussion of those English proficiency characteristics that were easier, more difficult, or distinguishing of minimally competent versus incompetent skills for target candidates. This was done so that the individual panelists could more clearly develop in their own mind what is the minimum English proficiency threshold for safe practice. By asking the entire group to provide and discuss these characteristics, every panelist in the group could consider different perspectives and rationales regarding what that minimum threshold should be. The particular examples and rationales were interesting, but they were given so quickly that they couldn't be sufficiently

recorded by the note-takers. Nevertheless, the process offered the panelists the opportunity to consider a variety of perspectives. This discussion occurred for each of the content domains that correspond to an IELTS module.

Prior to their operational ratings, panelists were given an opportunity to practice with the specific methods to ensure that they understood the task and how their judgments would be used. Two standard setting procedures were employed, a modified Angoff (Angoff, 1971; Impara & Plake, 1997) method for the Listening and Academic Reading subtests and a modified Analytical Judgment Method (Plake & Hambleton, 2000) for the Speaking and Academic Writing subtests. These methods are described briefly below.

Analytical Judgment Method. The Analytical Judgment method used is a modification of the method described by Plake and Hambleton (2000). This method entails asking panelists, in this case practicing nurses, to classify entry-level nurses' performance into defined categories. Classification is first at a broad level and then narrowed down to identify the performance that would likely be produced by a target entry-level nurse. This method was used for the Writing and Speaking modules of the IELTS.

Yes/No variation of the Angoff Method. The Yes/No Variation of the Angoff (1971) method (Impara and Plake, 1997) entailed asking nurses to examine each item on the test and estimate how a typical borderline "Minimally Competent" entry-level nurse for whom English is a second language will perform on that item. For the IELTS, panelists were asked to conceptualize a specific minimally competent nurse with whom they had worked or supervised. Keeping this entry-level nurse in mind, they were directed to indicate, for each item, whether the entry-level nurse they had in mind would answer the item correctly or not (Right or Wrong). This was done for the multiple-choice, short answer, and completion items the nurses rated. After an initial rating, actual performance data (proportion answering each item correctly) from a representative sample of over 8,000[6] IELTS test takers was provided to the panelists. After seeing the data, the panelists were asked to make a second estimate of whether the "Minimally Competent" entry-level nurse would answer correctly or not. The second estimate could be either the same or different from their first estimate. These data provide a reality check to ensure that expected performance is not set either unrealistically high or low because the nurse has misjudged how hard or easy the item actually is. The cut score is based on the second estimate and is calculated by summing, for each panelist, the number of "Right" items and then averaging those values across the panelists. This method was used for the Listening and Academic Reading modules of the IELTS.

Raters

The composition (qualifications, representativeness, and number) of the standard-setting panel is a major source of validity and credibility, as these are the experts who are recommending the passing score. Raymond and Reid (2001) advise that participants in a standard setting panel should: have subject matter expertise combined with a good understanding of the examinee population and their instructional environment; be able to estimate item difficulty, or receive training if necessary; appreciate the consequences of standard setting; and collectively represent communities of interest. Jaeger (1991) suggests that 13 panelists may be sufficient, while Cizek recommends using as many participants as resources and practical constraints will allow (Cizek, 1996). Given the context of the test (safe, entry-level nursing practice), the intended candidate population, and NCSBN's expectation that the IELTS passing score will be applied nationally, a panel of 28 experts was identified. These experts were believed to represent diverse, yet relevant professional perspectives.

---

[6] Test takers in this sample include candidates for professional positions and college admissions in addition to nurse candidates. It wasn't possible to isolate a sample of nurses; however the panelists were made aware of this limitation and were instructed to consider that limitation in any adjustments that they made.

The group was composed of: 10 licensed and practicing nurses with a range of experience who speak English as a second language and have taken the IELTS exam; 7 licensed and practicing nurses with a range of experience who were educated in the United States, who are currently working with clients who speak the most frequently identified non-English languages in the United States; 6 clinical supervisors of licensed and practicing nurses with experience working with entry-level nurses for whom English is not their first or primary language; 3 nursing regulators; 1 nursing educator; and 1 public member. The panelists represented all four NCSBN geographic regions (west, midwest, southeast, and northeast) and the major specialty/practice settings identified in NCSBN's most recent practice analysis (Smith & Crawford, 2003a; Smith & Crawford, 2003b).

Results

Panel Recommendations

Listening. The recommended cut score for the Listening module is based on the panelists' Yes/No ratings (Impara & Plake, 1997) from the selected and constructed response items. All items in this module were scored dichotomously (right/wrong). Providing actual performance data between rounds one and two appeared to have little influence on the panelists second ratings which suggests that either the data confirmed the panlists' ideas of how difficult the item was or it failed to persuade them to modify their rating. The cut scores and associated ranges within which the final cut score might be set as a result of using the Yes/No method are shown in Table 4. Panelists received their individual ratings and feedback in raw scores; however, the converted scale score in the IELTS's band score (0-9) metric is also reported. If the cut score were set at the average final value across the panel, it would be 29 out of a possible total of 40 (SD = 4). The impact of this cut score would be that 67% of the examinees that took the IELTS Listening module would be classified as being Incompetent. Note this does not provide specific information on the percent of nursing candidates because this sample included all examinees in the examinee pool. If the cut score were set at one standard deviation below the average cut score (25) the impact would be that 49.5% of examinees would be classified as being Incompetent.

Academic Reading. Because the items in this module were also scored dichotomously, the recommended cut score for the Academic Reading module is based on the Yes/No ratings from the selected and constructed response items. Impact data may have influenced panelists between rounds one and two as the second round resulted in a lower score mean and a higher standard deviation. Panelists received their individual ratings and feedback in raw scores; however, the converted band score is also reported in Table 5. The cut scores and associated ranges within which the final cut score might be set as a result of using the Yes/No method are also shown. If the cut score were set at the average final value across the panel, it would be 24 out of a possible total of 40 (SD = 5.45). The impact of this cut score would be that 62.7% of the examinees who took the IELTS Academic Reading module would be classified as being Incompetent. Again, this does not provide specific information on the percent of nursing candidates because this sample included all examinees. If the cut score were set at one standard deviation below the average cut score (19) the impact would be that 42.2% of the examinees would be classified as being Incompetent.

Speaking. The recommended cut score using the Analytical Judgment (Hambleton & Plake, 2000) method for the Speaking module is shown in Table 6. After viewing a series of video clips from the first section of the speaking module that represented a range of performances from international candidates, panelists were asked to identify the two speaking performances that represented the worst of the Competent performance and the best of the Incompetent performance. Next, the panelists were asked to repeat this for video clips from the second section. These four performances created a "bracket" around the panelist's recommendation of the cut score between the Minimally Competent and Incompetent candidates. Those averaged band

scores resulted in a first round cut score of 5.5 (SD = 0.7) on a scale of 0-9. Panelists were given feedback data on their individual cut score and the mean of the panelists' cut scores. They were then given the opportunity to change the performances they identified as the worst of the Competent performances and the best of the Incompetent performances. This resulted in the second round cut score of 5.6 (SD = 0.67). Table 6 also includes the range of results for 2 standard deviations above and below the recommended second round cut score.

Academic Writing. The Academic Writing module also used the Analytical Judgment method because the writing performances were scored using a multi-point (polytomous) scoring rubric. Table 7 shows the results for Task 1 of the Academic Writing module. Because there was a larger pool of potential performances from which panelists could select, panelists were asked to identify three writing performances that were the worst of the Competent performances and the three best of the Incompetent performances. Those averaged scores resulted in a round one cut score of 5.2 (SD = 0.53). Panelists were given feedback data on their individual cut score and the mean of the panelists' cut scores. They were then given the opportunity to change the performances they identified as the worst of the Competent performances and the best of the Incompetent performances. This resulted in the round two cut score of 5.3 (SD = 0.49).

The cut scores for Task 2 are also shown in Table 7 and were set using the same method as for Task 1 of the Academic Writing module. Again, panelists were asked to identify three writing performances that were the worst of the Competent performances and the three best of the Incompetent performances. Those averaged scores resulted in a round one cut score of 5.4 (SD = 0.34). Panelists were given feedback data on their individual cut score and the mean of the panelists' cut scores. They were then given the opportunity to change the performances they identified as the worst of the Competent performances and the best of the Incompetent performances. This resulted in the round two cut score of 5.4 (SD = 0.35).

In the operational IELTS test, the two individual scores taken from Task 1 and Task 2 in the Academic (and General Training) Writing module of the IELTS are combined into one overall band score. The overall band score for writing is computed by applying a 1/3 weight to the band score from Task 1 and a 2/3 weight from the band score from Task 2. A conversion grid is provided by IELTS to transform the two independent band scores from the writing tasks into one band score for the Writing module. For the purpose of this standard setting, the final band score was calculated using the panelists' cut score for each task and multiplying it by the weighting for each task and summing those scores. Table 7 also includes this combined final recommended band score. Pooled standard deviations were calculated to determine the amount of error present in combining the two cut scores for the Academic Writing module.

Summary of Panel Recommendations. The panel's standard setting judgments produced the following recommendations. For both Listening and Academic Reading modules, the average band score recommended was 6.5. These averages did not change from the first round to the second. For the Speaking module, the average band score was 5.5 in the first round and 5.6 in the second round. However on the Speaking module, candidate scores can only be whole numbers. Therefore, this average has to be interpreted as either a 5 or a 6. For the Academic Writing module (Task 1 and Task 2 combined), the average band score was 5.3 in the first round and 5.4 in the second. The Academic Writing module also does not permit half point scores; therefore, the recommended score must be interpreted as either a 5 or a 6.

Confidence in panelist ratings. Because appropriate training in the proper procedures combined with awareness of the consequences is recognized as vital in standard setting exercises (Raymond and Reid, 2001), the facilitator went to great lengths to ensure that the panelists understood what they were to do and why. To provide evidence of procedural validity, the panelists completed an evaluation form at the conclusion of the standard setting workshop. The evaluation form included questions related to six parts of the process. Part 1 focused on the orientation and training; Parts 2 and 3 focused on the levels of confidence, comfort, and length of

time for Rounds 1 and 2 of the Analytical Judgments; Parts 4 and 5 focused on Rounds 1 and 2 of the Yes/No ratings and on the levels of confidence and comfort in making the performance estimates and on the amount of time allowed to make the ratings; Part 6 assessed the overall workshop quality. An open-ended item asking about recommended changes that might be made to improve the workshop or make future workshops run more smoothly was also included at the end of Part 6. Results from the evaluation suggest that panelists were generally positive about their understanding of the process and confident in their judgments about the target candidate.

Discussion

NCSBN's Examination Committee was charged with recommending, on behalf of NCSBN, a passing standard for the IELTS examination that nurses should meet or exceed to be considered adequately proficient in English to use their nursing skills. In addition to the recommendations from the standard setting panel, the Examination Committee also considered the following information.

Existing Standards for Nurses. IELTS standards for the Commission on Graduates of Foreign Nursing Schools (CGFNS) certification and Visa screening already exist. Other countries and other professions have also set standards (minimum scores) on the IELTS scale for licensing purposes. Table 8 provides some detail regarding these standards; however a quick glance at the table indicates that these groups generally regard the minimum level of English proficiency as measured by IELTS to somewhere between 5.5 and 7.5.

Impact. To anticipate the potential impact of a particular passing standard, Table 9 is presented. However, the data in this table must be interpreted cautiously. The cumulative percentages are conceptually a module-specific fail rate prediction by band score for candidates taking the Academic version of the test in 2003 and 2004 for purposes of employment, professional registration, or immigration. Given that an examinee must pass all the requirements (if a standard is specified by module) or a single aggregated requirement (an overall score), it seems reasonable to assume that the actual fail rate for nurse examinees will not necessarily match what is predicted in Table 9. Although these data give us some idea about the impact among a group of candidates who took the Academic version of IELTS, we do not fully know the extent to which this group's characteristics overlap with the nursing candidate population. Thus, this group may not be typical of IELTS examinees that wish to work in the US as a nurse. How a candidate performs on the individual modules adds another layer of complexity. Scores across modules are going to be related (because they all measure an aspect of English Proficiency), but it is certainly possible to have some variation across scores as well.

Additional policy considerations

As a further illustration of the policy nature of the standard setting process, this section describes the different decision scoring options that the Examination Committee considered when advising NCSBN. As mentioned above, the committee considered the available information from this panel in combination with information gathered from other nursing programs as well as other professions that have incorporated an English proficiency test into their licensure process (See Table 8). The committee's discussion focused on the contribution of English proficiency to the necessary entry-level knowledge and skills for the nursing profession. They also considered the characteristics of the IELTS band score descriptors (See Table 2) in their discussion.

The Examination Committee then discussed the nature of the scoring decision and the implications of each option. The first option would be a *compensatory* scoring decision whereby the pass/fail decision would be based on the candidate's total test score across the IELTS modules. A compensatory approach may be appropriate when there is evidence of a single underlying construct, if performance on the separate modules is at least moderately correlated with one another, and if the policy can reasonably allow candidates to compensate low performance in one domain with higher performance in another domain. Another factor that is critical to the health professions is the policy consideration of Type I (incompetent passers) and Type II (competent failers) errors and the impact that these decision errors may have on the

public. When there is a greater public risk, more Type II errors (competent failers) may be tolerated.

The second option the committee considered was a *conjunctive* approach. This strategy would establish a passing score for each module and require candidates to meet the passing score within each module to fully pass the test. This decision scoring model may be used when there are distinct, uncorrelated domains or when policy bodies view each domain as equally important to the licensure decision such that minimum competency is required for each one. Because it has more passing scores (one for each module) than the compensatory approach (one for the total test score), the passing rate for a conjunctive decision will be at least equal to the compensatory decision, but more than likely, lower. This effectively reduces the number of incompetent passers, but potentially increases the number of competent failers. This is why the use of the scores in the context of the nursing profession is important to define at the outset.

A third option that the committee considered resulted in their ultimate policy decision. This option could be characterized as a *hybrid* of the compensatory and conjunctive approaches. The hybrid approach establishes an overall passing score based on a compensatory decision (e.g., a band score of 6.5 for the overall total score), but establishes minimum values for each module (e.g., a band score of 6.0) that is slightly below the overall requirement. This approach acknowledges the error that is inherent in all cognitive measurement and allows for some compensating skills to benefit the candidate. However, it precludes a candidate from performing miserably in one module (e.g., Speaking) and offsetting it with an exceptional performance in another module (e.g., Reading). This approach will result in a passing rate that falls somewhere between the fully compensatory and fully conjunctive decision scoring strategies.

Based on their deliberations, the Examination Committee approved a passing score that would require a band score result of 6.5 overall with a minimum of 6.0 in any one module as the NCSBN recommended passing standard for the IELTS examination. The committee also recommended that international nurse-candidates should meet or exceed this standard before they would be issued a license. The NCSBN Board of Directors approved the committee's recommendation. In August 2005, this standard was presented to the NCSBN Delegate Assembly, which permitted the NCSBN recommended IELTS passing standard for nurse licensure to stand. This recommended passing standard was subsequently communicated to the public via press release. Given the combination of measurement and policy considerations, this effort was accomplished over a 12-15 month timeframe.

Implementation

It is important to note that NCSBN has worked to provide a recommended standard that is legally defensible and serves to protect the public as a benefit to its membership, the boards of nursing. The act of licensing professions is an authority that rests with the state, not with a not-for-profit organization like NCSBN. Only the state has the authority to define the requirements for licensure. The question now is, how many of the boards of nursing will use this standard as a legal requirement for licensure? In a similar vein, the adoption of these standards for U.S. visa screening purposes is also of interest. Because the adoption of this standard rests entirely with the individual state, NCSBN has little more to do with implementing the standard, other than answering questions and providing documentation about how the standard was set.

Conclusions

There are a number of tests that seek to measure international students' or candidates' English language proficiency. Since tests like the IELTS and TOEFL both claim to measure English proficiency, and NCSBN had adopted both of these tests, one might wonder why it is necessary to set a new standard on IELTS? A more straightforward approach might appear to be an attempt to equate the TOEFL standard onto the IELTS scale. However, attempting to directly link these or other English proficiency tests would likely be tenuous for two reasons. First, the construct of English proficiency may be represented differently across the two examinations, in part due to the differences in performance demands. Specifically, the way the English proficiency

construct is operationalized has until recently differed substantially across the two tests, both in terms of the skills tested and the test methods used. Furthermore, the TOEFL reflects English as spoken in North America, while IELTS has been developed with an international perspective on the use of English, in which North American usage is only one dimension. Such differences in construct definition and representation make it difficult to undertake equating studies and to establish meaningful equivalence between the two tests.

Second, in conducting an equating study, one typically links two tests using common items, common people, or equivalent groups of people. It is doubtful that the TOEFL and IELTS have any common items. In fact, the existence of common items might indicate that one party was infringing upon the copyright of the other. Even if NCSBN could assemble a cohort of people who have taken both examinations, implementing such a design would still be vulnerable to differences in the construct. Similarly, equating using equivalent groups would produce tenuous results unless it was really possible to randomly assign candidates to one test or the other. NCSBN does not have the authority to randomly assign candidates to take particular tests nor does it have evidence that candidates who take the IELTS are randomly equivalent in their backgrounds or abilities to those candidates that take the TOEFL.

The utility of having a common construct that would permit conversion of one scale to another is obvious because licensing boards want the assurance that the meaning of the passing score is approximately the same regardless of the test. In this situation, examinees that take one test would more readily be able to be compared to examinees that have taken another test. The recent inclusion of a speaking component on the TOEFL (ETS, 2005) means that the IELTS and TOEFL tests may now be more closely aligned in format and content than they were previously. A study examining this question may make it easier for test users (e.g., licensing boards) to make meaningful and useful comparisons between two tests which are being used in similar ways.

As licensure testing programs face the challenge of measuring the language competence of internationally trained candidates as well as their professional competence, studies such as the one described here will be necessary to inform policymakers and the testing community. This article described the processes used to derive an English proficiency passing standard specifically for licensing nurses in the US using IELTS. This article provides an illustration of this process and the policy considerations that NCSBN used when developing a nursing-specific passing standard on the IELTS that states could consider in their licensure decisions for internationally trained candidates. The information from this study was considered by NCSBN's Examination Committee in conjunction with other relevant information to produce a legally defensible passing standard.

Kane (1998) describes the setting of cut scores as 'an exercise in policymaking that can be informed by expert judgment and input from stakeholders but is not determined by any empirical investigation' (p.137). Zieky (2001) acknowledges that because cut scores are dependent on the value judgments of those involved in making them, they will always be controversial and subject to challenge. Similarly, the adoption by licensure testing programs of language assessment tools which are less, rather than more, 'occupation-specific' will continue to be the subject of debate. Despite their limitations, studies such as the one described in this paper will surely continue to play an important role in setting policy standards in contexts where issues of language proficiency interface with those of professional competence. In today's globalized world where labor flow between countries continues to increase, we believe the matters of policy and practice discussed here have international relevance, and it is our hope that this article will be of interest and value to other testing programs, language testers and policymakers facing similar challenges worldwide.

References

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

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement*, (2nd ed.), p. 508-600. Washington, DC: American Council on Education.

British Council, IELTS IDP Australia, & University of Cambridge ESOL Examinations. (2003). *IELTS International English Language Testing System: Handbook*. Cambridge, UK: author.

Camilli, G., Cizek, G.J. & Lugg, C.A. (2001). Psychometric theory and the validation of performance standards: history and future perspectives. In Cizek, G. J. [Ed.] (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cizek, G.J (1993). Reconsidering standards and criteria. *Journal of Educational Measurement, 30 (2),* p. 93-106.

Cizek, G.J. (1996). Setting passing scores. *Educational Measurement: Issues and Practice, 15* (2), 20-31.

Cizek, G. J. [Ed.] (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cizek, G. J. & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage Publications.

Educational Testing Service. (2003). *TOEFL 2003-04 Information bulletin for computer-based and paper-based testing.* Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2005). *2005-06 Test of English as a Foreign Language(TM) Information and Registration Bulletin for Internet-based Testing (iBT).* Princeton, NJ: Educational Testing Service.

Hambleton, R. K. & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.). Westport, CT: American Council on Education and Praeger Publishers.

Impara, J. C. & Plake, B. S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement, 34*(4), 353-366.

Jaeger, R. J. (1991). Selection of judges for standard setting. *Educational Measurement: Issues and Practice, 10*(2), 3-6, 10, 14.

Kaftandjieva, F. (2004). Standard setting, Section B in the *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the Common European Framework of Reference for Languages: learning, teaching and assessment*. Retrieved 16 July, 2006 from http://www.coe.int/T/DG4/Portfolio/?L=E&M=/documents intro/Reference supplement intr.html

Kane, M. (1998). Choosing between examinee-centred and test-centered standard-setting methods. *Educational Assessment, 5 (3),* 129-145.

O'Neill, T. R. (2004). The minimum proficiency in English for entry-level nurses: TOEFL®. *NCLEX Psychometric Research Brief, volume 1*. Retrieved June 5, 2006 from http://www.ncsbn.org/pdfs/TOEFL_Research_Brief_vol_1.pdf.

O'Neill, T. R., Marks, C., & Wendt, A. (2005). Recommending a minimum English proficiency standard for entry-level nursing. *JONA's Healthcare Law, Ethics, and Regulation, 7 (2),* 56-58.

O'Neill, T. R., Marks, C., & Reynolds, M. (2005). Re-evaluating the NCLEX-RN® passing standard. *Journal of Nursing Measurement, 13 (2,),* 129-146.

O'Neill, T. R., Tannenbaum, R., & Tiffen, J. (2005). Recommending a minimum English proficiency standard for entry-level nursing. *Journal of Nursing Measurement, 13 (2), 129-146*.

Plake, B. S., & Hambleton, R. K. (2000). A standard-setting method designed for complex performance assessments: Categorical assignments of student work. *Educational Assessment, 6*, 197-215.

Raymond, M.R., & Reid, J.B. (2001). Who made thee a judge? Selecting and training participants for standard setting. In Cizek, G. J. [Ed.] (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Smith, J. E., & Crawford, L. H. (2003a). *2002 RN Practice Analysis: Linking the NCLEX-RN Examination to Practice.* Chicago: National Council of State Boards of Nursing.

Smith, J. E., & Crawford, L. H. (2003b). *2003 LPN/VN Practice Analysis: Linking the NCLEX-PN Examination to Practice.* Chicago: National Council of State Boards of Nursing.

Zieky, M. J. (2001). So much has changed: how the setting of cutscores has evolved since the 1980s. In Cizek, G. J. [Ed.] (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Table 1. Internationally-educated, First-time NCLEX-RN Examinees

Year	Count
2001	8,613
2002	12,762
2003	16,490
2004	18,285
2005	17,980
2006	25,908

Table 2. IELTS Overall Band Scores
Band 9 – Expert User
Has fully operational command of the language: appropriate, accurate and fluent with complete understanding.
Band 8 – Very Good User
Has fully operational command of the language with only occasional and unsystematic inaccuracies and inappropriacies. Misunderstandings may occur in unfamiliar situations. Handles complex detailed argumentation well.
Band 7 – Good User
Has operational command of the language, though with occasional inaccuracies, inappropriacies, and misunderstandings in some situations. Generally handles complex language well and understands detailed reasoning.
Band 6 – Competent User
Has generally effective command of the language despite some inaccuracies, inappropriacies, and misunderstandings. Can use and understand fairly complex language, particularly in familiar situations.
Band 5 – Modest User
Has partial command of the language, coping with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle most communication in own field.
Band 4 – Limited User
Basic competence is limited to familiar situations. Has frequent problems in understanding and expression. Is not able to use complex language.
Band 3 – Extremely Limited User
Conveys and understands only general meaning in very familiar situations. Frequent breakdowns in communication occur.
Band 2 – Intermittent User
No real communication is possible except for the most basic information using isolated or short formulae in familiar situations and to meet immediate needs. Has great difficulty in understanding spoken and written English.
Band 1 – Non User
Essentially has no ability to use the language beyond possibly a few isolated words.
Band 0 – Did not attempt the test
No assessable information provided

Table 3. Panel's description of English difficulty levels across activities

	Listening	Speaking	Reading	Writing
Easier	Using non-verbal clues Common terminology	Verb tense Academically correct English	Nurse notes History / physical record Medication list	Routine nursing documentation Familiar terminology and context
Harder	Non-standard usage, jargon, and abbreviations	Using colloquialisms / slang Multiple meanings for the same word Culturally-specific English usage	Doctor's orders Medical / technical terms Consultant reports Diagnostic reports	Documenting using non-routine words Non routine tasks (e.g., patient complaints, legal documentation, etc.)
Distinguish between incompetent and minimally competent	Fine distinctions among common words that are critical to practice. Context cued distinctions Speed of comprehension	Appropriate use of technical language (jargon) Questioning to clarify meaning Sensitivity to the message being communicated Fluency, ability to generate more developed sentences	Comprehension / understanding Speed of reading Word recognition Grasp of grammar Familiarity with the context	Vocabulary and word choice Accuracy of words. Spelling Communicating and interpreting events (sequencing, organizing) Technically correct grammar Maintaining relevance (staying on target) Translating patient's words into text.

Table 4.
Recommended cut scores for the Listening module

	Score (SD) (out of 40)	Impact[a] (% below)	Band Score
Round 1 Mean Score (SD)	29 (4)	67.0%	6.5
Round 2 Mean Score (SD)	29 (4)	67.0%	6.5
Round 2 minus 2 SD	21	33.3%	5.5
Round 2 minus 1 SD	25	49.5%	6.0
Round 2 plus 1 SD	33	83.5%	7.5
Round 2 plus 2 SD	37	95.9%	8.5

[a] Impact (% below) refers to the cohort of all IELTS examinees that would have failed based upon this standard, assuming a population similar to the September 2003-August 2004 cohort of examinees who indicated that their destination was the United States and that their reason for taking the examination was either professional registration, medical council employment, or immigration.

Table 5.
Recommended cut scores for the Academic Reading module

	Score (SD) (out of 40)	Impact[a] (% below)	Band Score
Round 1 Mean Score (SD)	26 (3.9)	69.6%	6.5
Round 2 Mean Score (SD)	24 (5.5)	62.7%	6.5
Round 2 minus 2 SD	13	14.6%	5.0
Round 2 minus 1 SD	19	42.2%	5.5
Round 2 plus 1 SD	29	78.6%	7.0
Round 2 plus 2 SD	35	93.4%	8.0

[a] Impact (% below) refers to the cohort of all IELTS examinees that would have failed based upon this standard, assuming a population similar to the September 2003-August 2004 cohort of examinees who indicated that their destination was the United States and that their reason for taking the examination was either professional registration, medical council employment, or immigration.

Table 6.
Recommended cut scores for the speaking module.

	Band Score
Round 1 Mean Score (SD)	5.5 (0.70)
Round 2 Mean Score (SD)	5.6 (0.67)
Round 2 minus 2 SD	4.5
Round 2 minus 1 SD	5.0
Round 2 plus 1 SD	6.5
Round 2 plus 2 SD	7.0

Table 7.

Recommended cut-scores for the Writing module

Writing Task 1	Score[a]
1st round mean score (SD)	5.2 (0.53)
2nd round mean score (SD)	5.3 (0.49)
Writing Task 2	
1st round mean score (SD)	5.4 (0.34)
2nd round mean score (SD)	5.4 (0.35)
Task 1 & Task 2 (pooled)	
1st round mean score (SD)	5.3 (0.63)
2nd round mean score (SD)	5.4 (0.60)
2nd round minus 2 SD	4.0
2nd round minus 1 SD	5.0
2nd round plus 1 SD	6.0
2nd round plus 2 SD	6.5

[a]IELTS band scores are reported to the nearest half point; however means and standard deviations are reported to the nearest tenth to minimize rounding error when deviations from the mean are computed.

Table 8.
Other IELTS passing standard for other professions and countries

Country	Purpose	IELTS Band Score
Nurses		
US	Visa screen or CGFNS certification for RNs	Band 6.5 overall (min 7 in Speaking)
US	Visa screen or CGFNS certification for PNs or VNs	Band 6.0 overall (min 7 in Speaking).
Australia	Australian Nursing Council	Band 7.0 overall (min 6.5 in Listening and Reading, min 7 in Writing and Speaking)
New Zealand	Nursing Council of New Zealand	Band 7.0 overall (min 6.5 in Listening and Reading, min 7 in Writing and Speaking)
Ireland	An Bord Altranais (Irish Nursing Board)	Band 6.5 overall (min 6.0 in Writing, min 5.5 in Reading and Listening)
Canada	College of Nurses of Ontario (It is a Board of Nursing)	Band 6.5 overall (min 6.5 in each module)
Canada	Registered Nurses Association of British Columbia (It is a Board of Nursing)	Band 6.5 overall (min 7 in Speaking) – 6.0 in Reading, Writing and Listening
UK	Nursing & Midwifery Council	Band 6.5 overall (min 5.5 in each module) – 5.5 in Listening and Reading, 6.0 in Writing and Speaking
Other Professions		
US	American Association of Veterinary State Boards (PAVE Program)	Band 7.0 overall
US	American Veterinary Medical Association (ECFVG Program, DVMs from foreign/non-accredited schools))	Band 6.5 overall (min 6 in Writing, 6.5 in Listening, 7 in Speaking)
US	Visa screening for Medical Technologists and Clinical Laboratory Scientists.	Band 6.5 overall (min 7 in Speaking)
US	Visa screening for Medical Technicians and Clinical Laboratory Technicians.	Band 6.0 overall (min 7 in Speaking)
US	Visa screening for Physicians Assistants	Band 6.5 overall (min 7 in Speaking)
Canada	Ontario College of Pharmacists	Band 7.0 overall (min 6 in each module)
UK	General Medical Council	Min of 7 in each module – without PLAB (Professional & Linguistics Assessment Board) Band 7 overall. Min 7 in Speaking, 6 in Reading, Writing and Listening – with PLAB
New Zealand	Medical Council of New Zealand	Band 7.5 overall
Australia	Australian Medical Council	Band 7 overall

Most of this information comes from an IELTS Band Score Requirements 2004, an IELTS fact sheet, and some independent verification via web or phone.

Table 9.

Cumulative Percentages for Academic candidates in 2003-2004 by Band Score and Test Component.

Band Score	Academic Reading	Academic Writing	Listening	Speaking
1	0.02	0.00	0.01	0.00
2	0.03	0.03	0.01	0.00
3	0.05	0.09	0.09	0.05
3.5	0.11		0.23	
4	0.64	1.22	0.68	0.65
4.5	3.75		2.94	
5	14.53	14.15	9.84	8.21
5.5	36.09		26.29	
6	62.01	60.25	50.05	46.91
6.5	80.55		72.68	
7	91.23	95.23	87.51	91.77
7.5	96.08		94.74	
8	98.21	99.59	98.03	98.29
8.5	99.30		99.07	
9	100%	100%	100%	100%

These figures relate to all IELTS test-takers taking the Academic Modules between Sept 2003 and Aug 2004 who gave the US as their "country of application" (i.e. destination) and gave their reason(s) for taking IELTS as professional registration, medical council employment, or immigration.

Reading and Listening scores are reported using half-bands, while Writing and Speaking scores are reported using whole bands only.