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

"This is a preprint of an article whose final and definitive form has been published

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

]TJ0.0sce. dTr1 Tc 0.0004 Tw4BDC BT/rotects t[(ti(em).9453 -use of e.02 0uepeat/Ppan 6 Tw T[(program)8(s, thch 8

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

“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

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 .104 -1.15ewp0vd9e

The group was composed of: 10 licensed and practicing nurses with a range of experience

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.,

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

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

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

	Listening	Speaking	Reading	Writing
Easier	Using non-verbal clues			

Commf/TT0 0 1 4.0784 Tw 0 10.02 -u9 Tm[(Easi)3/P <MCID 8 BDC odogy ac0.0029 Tc -0.0014 Tw 0 1984 72 Tm( )TjETEMC /P <C)55-u9 Nu0 1rsen0 1o/P tT

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