

State boards of nursing are charged with protecting the public through the regulation of nursing in their jurisdiction. One of the ways in which they perform that function is by setting and enforcing prerequisite conditions for getting a license. Often these prerequisites are related to education, experience, and demonstrating that they are at least minimally competent

Using an underlying principle of public safety, the minimum standard was intended to reflect the level of English language proficiency believed necessary for entry-level nurses to be able to perform important nursing responsibilities safely and effectively. It is recommended that internationally educated nurse-candidates meet or exceed this standard before they are issued a license. It is important to note that the standard was intended to reflect the minimum level of English proficiency necessary for safe and effective-

sionals who participate in its definition and adoption, and different professionals may hold different sets of values. Its determination may be informed by empirical information or data, but ultimately, the passing standard is a judgment-based decision.

Regardless of one's theoretical perspective, the standard used to classify examinees must not be made in an arbitrary and capricious manner. Furthermore, the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999), recommend that the rationale and procedures used to set the standard be clearly documented. This includes 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 report documents these aspects of the standard setting process.

METHODS

TOEFL examination

The TOEFL is an examination designed to assess English language ability in examinees for whom English is not their native language. With the exception of the writing section, all items are dichotomously scored and use a Selected Response (SR) type item format (essentially multiple-choice-question format, although the mechanism for selection can vary a bit and multiple responses are required in some instances). The writing sample is a single prompt that is scored 0-6. Three section-level subscores (0 to 30 points each) and a total TOEFL score (0 to 300 points) are reported. The total TOEFL score is the average of the three section scores multiplied by 10. The three sections are: (1) Listening Comprehension, (2) Structure and Written Expression, and (3) Reading Comprehension.

Standard Setting Methods

Two standard setting procedures, the *Simulated Minimally Competent Candidate* (SMCC) method for

the selected response format items and the *Examinee Paper Selection Method* for the essay portion, were combined to produce passing standard recommendations for each panelist. Because some sections of the TOEFL are adaptive and the test uses item response theory (IRT) to equate all examinee's performances to a common scale, it was desirable to use a standard setting procedure that was congruent with adaptive testing and IRT. The SMCC method is such a procedure. This method essentially asks each panelist to respond to a sample of items the way they imagine a minimally competent examinee would. Based upon those responses, a score is computed for the panelist that should represent the panelist's notion of minimal competence. If the items are

that the minimally competent candidate would perform somewhere between two adjacent exemplars. This is consistent with the actual scoring process because two raters grade each essay and the average rating is used (Educational Testing Service, 2003). A conversion table was used to combine the selected rating with the partial subtest score for Structure & Written Expression that was generated using the SMCC procedure. The SMCC procedure and the *Examinee Paper Selection* method each contributed approximately 50% of the Structure/ Written Expression subtest.

Finally, there were three subtest scores for each panelist that could have ranged from 0 to 30. These subtest scores were combined into a total score (0 to 300) by summing the three section scaled scores and multiplying this sum by ten-thirds, effectively allowing each section scaled score to contribute equally to the total scaled score.

Adaptive Testing

Reading was administered as a fixed form test, but the Listening Comprehension and Structure sections were administered adaptively. That is, the difficulty level of an item presented to a candidate is dependent on the candidate's response to the immediate previous item and to the other previous items. A correct response to an item, for example, is followed by an item of greater difficulty; an incorrect response is followed by an item of lesser difficulty. In this way, a candidate receives a set of items maximally tailored to his or her overall ability in each of the two adaptive sections.

PowerPrep®

Each panelist was provided with a laptop computer that was preloaded with TOEFL PowerPrep software. PowerPrep contains two full-length, computer-adaptive editions of the TOEFL, drawing upon a pool of more than 1,200 items. The software does not provide a final score for the Structure & Written

Expression section, but instead it produces the lower bound of the Structure score, which essentially assumes that zero points were earned on the essay. Panelists were to combine this Structure score with their essay score through the use of a conversion table to produce a single score for the Structure & Written Expression section. For each section, the panelist's ability estimate was translated to a scaled score that could range from 0 to 30. Finally, a total score (0 to 300) for the panelist was obtained by summing the three section scaled scores and multiplying this sum by ten-thirds, effectively allowing each section scaled score to contribute equally to the total scaled score.

Selection of Raters

The composition (number, representativeness, and qualifications) of the standard-setting panel was a crucial element in establishing the validity and credibility of the standard. Twenty-five experts served on the standard-setting panel (Table 1). The panel, as a group, was intentionally made ethnically and linguistically diverse. Applicants were grouped by the following categories a) having previously taken the TOEFL exam, b) working with clients who speak languages other than English, c) supervising nurses who speak languages other than English, or d) working as nursing regulators. NCSBN further sorted applicants by selecting candidates from each of the most commonly spoken non-English languages in the U.S., and selecting representatives from all four NCSBN geographic regions. These experts, all female, were recruited by NCSBN to represent a range of professional perspectives and experiences. Collectively, 18 jurisdictions were represented on the panel: AK, CA, DC, FL, GA, HI, IA, IL, KS, LA, MA, MN, NJ, NC, OH, OR, TX, and VA.

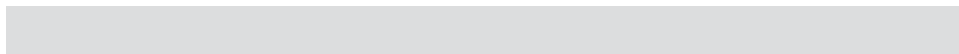
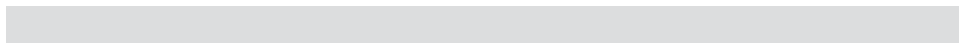
Panelist Orientation and Training

The panelists were first provided with an overview of the goals and purpose of the study. It was explained

THE READABILITY

that a passing score was meant to reflect the level of English language proficiency necessary for entry-level nurses to perform important nursing tasks safely and effectively. It was clarified that the passing score was not the level of English language proficiency necessary to take the NCLEX examination—the focus of the study was on the job. Second, the panelists were led through an overview of the TOEFL computer-based test and the general process that was to be followed in arriving at the recommended passing score.

After the orientation, the panel was asked to identify the core tasks that all entry-level nurses needed to perform. It was important to agree on the scope of activity that was being considered before trying to assess how much English one needed to know to perform them. The list included: taking patient histories, conducting patient assessment, completing documentation, educating-training patients, taking orders, reporting, implementing safety practices, delegating, communicating, providing client service, and prioritizing responsibilities. This list was posted



to serve as a frame-of-reference for the rest of the exercise.

Next, the panelists were instructed to imagine a nurse candidate who was educated outside the United States and in a language other than English. Furthermore, this imaginary candidate was seeking to become an entry-level nurse in the U.S. and just barely possessed the English proficiency necessary to be safe and effective as a nurse. Panelists were reminded that the focus was not on the examinee's nursing knowledge or skill, but rather on their English language skills.

The panelists were instructed to take the TOEFL



- THE READABILITY OF NCLEX® EXAMINATIONS
 NCLEX® PSYCHOMETRIC RESEARCH BRIEF | VOL 1 | SEPTEMBER 2004

Table 4. First and Second Round Scores for All Panelists who have taken TOEFL (11 panelists)						
Panelist	Listening (0-30)	Structure Component (0-13)	Writing Component (0-6)	Combined Structure and Writing (0-30)	Reading (0-30)	Total (0-300)
ROUND 1 JUDGMENT						
P1	22	12	3.5	23	19	213.33
P4	24	9	5	25	23	240.00
P5	25	11	4	24	24	243.33
P7	16	6	4	19	22	190.00
P12	10	6	4	19	24	176.67
P14	23	9	4	22	19	213.33
P16	20	10	3.5	22	24	220.00
P20	22	9	4	22	21	216.67
P21	22	5	4	18	16	186.67
P22	26	13	3.5	24	21	236.67
P23	24	5	4.5	20	18	206.67
Mean (truncated)	21	8	4	21	21	213
Median (truncated)	22	9	4.	22	21	213
Standard Deviation	4.39	2.67	0.43	2.23	2.59	21.01
Minimum	10	5	3.5	18	16	176.67
Maximum	26	13	5	25	24	243.33
ROUND 2 JUDGMENT						
P1	20	8	4	21	20	203.33
P4	24	9	4	22	22	226.67
P5	22	10	4.5	24	21	223.33
P7	20	6	4	19	22	203.33
P12	15	7	4	20	24	196.67
P14	23	9	4.5	24	22	230.00
P16	20	10	4.5	24	24	226.67
P20	22	8	4	21	21	213.33
P21	22	9	4	22	22	220.00
P22	26	11	3.5	23	25	246.67
P23	22	10	3.5	22	20	213.33
Mean (truncated)	21	8	4	22	22	218
Median (truncated)	22	9	4	22	22	220
Standard Deviation	2.68	1.40	0.33	1.60	1.56	13.73
Minimum	15	6	3.5	19	20	196.67
Maximum	26	11	4.5	24	25	246.67

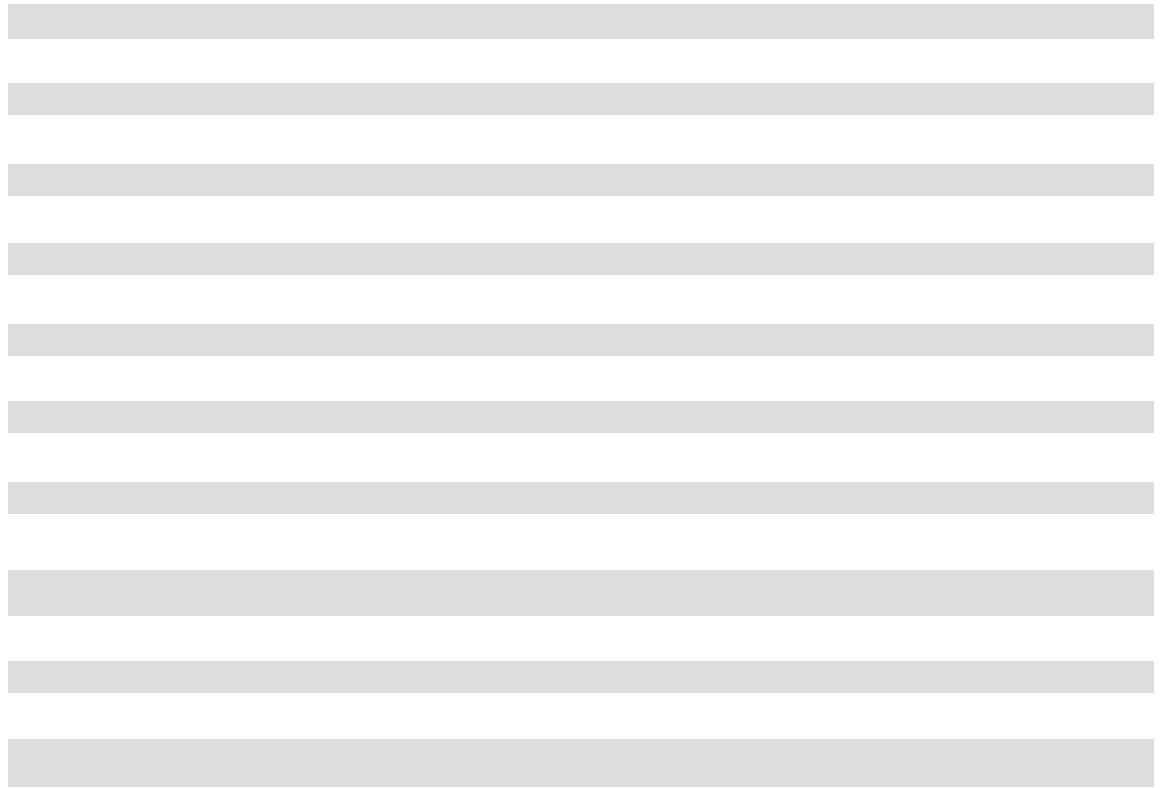


Table 5, continued

Panelist	Listening (0-30)	Structure Component (0-13)	Writing Component (0-6)	Combined Structure and Writing (0-30)	Reading (0-30)	Total (0-300)
ROUND 2 JUDGMENT						
P2	22	10	4	23	23	226.67
P3	26	9	5	25	26	256.67
P6	24	10	3.5	22	23	230.00
P8	24	8	4	21	20	216.67
P9	20	11	4.5	25	24	230.00
P10	18	8	3.5	20	22	200.00
P11	20	9	3.5	21	20	203.33
P13	23	11	4	24	23	233.33
P15	21	10	4.5	24	20	216.67
P17	22	10	4.5	24	24	233.33
P18	25	9	4	22	25	240.00
P19	21	8	3.5	20	19	200.00
P24	20	8	3.5	20	20	200.00
P25	22	11	4	24	25	236.67
Mean (truncated)	22	9	4	22	22	223
Median (truncated)	22	9	4	22	24	248.15

[Redacted text block]

[Redacted text block]

[Redacted text block]

[Redacted text block]

Foreign Nursing Schools (CGFNS, 2000) in their validity study, one could get a better idea of the typical distribution of English language proficiency for internationally educated nurses taking the TOEFL. Using only the people reported in that study who were in the 1999 or 2000 TOEFL cohort, an estimate for the population of nurses was derived ($\mu=237.5$, $s=19$)². Using this population, one would expect 82% ($z = -0.921$) of them to pass.

The Examination Committee considered the impact predictions and agreed that a standard of 220 on the TOEFL was appropriate to demonstrate the minimum degree of English proficiency necessary to be a safe and effective, entry-level nurse. Correspondingly, a score on of 560 on the paper version of the TOEFL would be considered equivalent.

DISCUSSION

The purpose of this study was to arrive at a recommended passing score on the TOEFL that represented the level of level of English language proficiency believed necessary to perform important entry-level nursing tasks safely and effectively. The Examination Committee was asked to make a policy decision after being informed with the appropriate types of information. The committee did this after giving consideration to a broad spectrum of information. As a result, there is now a recommended passing standard for entry-level nursing that can be supported by carefully documented and well-designed procedures.

Limitations

Typically, there are some shortcomings that are inherent in tests that are related to licensure and certification testing. Test developers are often restricted in the types of data that they can collect to verify the standard. In practice, boards only license or certify people that are believed to be competent. Were they believed to be incompetent, it would be unethical to license or certify them. Because these people come only from the upper end of the ability continuum, there are sampling problems related to attempting to establish the predictive validity of the standard. Therefore, this type of predictive validity is not normally demonstrated for certification and licensure tests.

Future Activities

Now that the standard has been set, the question is now how many of the boards of nursing will use this standard as a legal requirement for licensure? Also, the adoption of these standards for visa screening purposes is also of interest. Because the adoption and implementation of this standard rests with governmental entities, NCSBN's role is one of providing information and documentation about the standard. In the future, NCSBN intends to provide recommended standards for other English Proficiency examinations as well. This will provide boards of nursing and candidates with more choices in tests and test providers.

²The CGFNS TOEFL sample was based on the written examination, not the CAT examination. The written examination scores were converted to CAT scores via the following formula $CAT = (Written - 273.9) * 0.769$. This formula was based on a conversion table found on page 13 of the TOEFL 2003-04 Information Bulletin for Computer-based and Paper-based Testing.

REFERENCES

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

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

Commission on Graduates of Foreign Nursing Schools. (2000). Commission on Graduates of Foreign Nursing Schools Validity Study April 1999 through March 2000. Unpublished statistical report.

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

Hambleton, R.K., Jaeger, R.M., Plake, B.S., & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, 24, 355-366.

Hurtz, G.M., & Auerbach, M.A. (2003). A meta-analysis of the effects of modifications to the Angoff method on cutoff scores and judgment consensus. *Educational and Psychological Measurement*, 63, 584-601.

National Council of State Boards of Nursing, Inc. (2003). Profile of Member Boards 2003. Chicago, IL: National Council of State Boards of Nursing, Inc.

A special thanks to the members of the NCSBN Examination Committee who debated and deliberated to produce a well conceived standard. Also, thanks to Anne Wendt, RN, PhD, CAE, who recommended many of the specifications for the selection of panelist.



National Council of State Boards of Nursing
111 E. Wacker Drive, Suite 2900
Chicago, IL 60601-4277 USA
312.525.3600 | International Calls: +1.312.525.3600
Testing Services: 312.525.3750 | www.ncsbn.org