Re-Evaluating the NCLEX-RN®
Passing Standard

Thomas R. O'Neill, PhD
Casey M. Marks, PhD
Michelle Reynolds, MS
National Council of State Boards of Nursing
Chicago, IL

Setting passing standards is a critical component of the NCLEX® examination process. This research was conducted to provide sufficient information to the National Council of State Boards of Nursing's (NCSBN) Board of Directors to make a decision regarding the passing standard of the NCLEX-RN. This article illustrates the standard setting process that NCSBN uses. Surveys of educators and employers, a modified Angoff procedure, the Beuk compromise, and global assessments by content experts were methods used. The Rasch model and a presumed ability distribution were used as the framework to integrate these diverse perspectives regarding minimal competence. The revised passing standard was -0.28 logits. For many of the minimal competence estimates, the author did not have authorization to release the information. In those instances, estimates and results were fabricated to be similar to the actual results, yet different enough as to not disclose confidential information. The fabricated results are clearly marked. In conclusion, a variety of approaches, sources, and perspectives are necessary for the establishment of fair and appropriate standards on the NCLEX-RN.

Keywords: NCLEX; standard setting; passing standard; Angoff; NCSBN; licensure; Rasch Model

T he National Council of State Boards of Nursing, Inc. (NCSBN) is a not-for-profit organization that is composed of the jurisdictional boards of nursing in the United States and U.S. territories. NCSBN's mission is to provide leadership to advance regulatory excellence for public protection. One of the many ways that NCSBN fulfills this mission is by providing its members (boards of nursing) with a defensible method of assessing a candidate's competence. Specifically, NCSBN creates and administers two minimal competency examinations, the National Council Licensure Examination for Registered Nurses® (NCLEX-RN) and the National Council Licensure Examination for Practical Nurses® (NCLEX-PN). All boards of nursing that are members of NCSBN use the NCLEX as part of their licensing process. This article is intended to illustrate NCSBN's standard setting procedures and the conceptual framework behind the procedures using the April 2004 revision of the NCLEX-RN passing standard.

By virtue of being a licensing examination, the NCLEX-RN is a high-stakes examination. In a typical year, 120,000 NCLEX-RN examinations are administered. Therefore, when setting the passing standard, it is important to set it high enough to protect the public by being a barrier to incompetent nurses, yet also be low enough that competent nurses are not denied a license. This article discusses issues related to standard setting for a

national licensing test and describes how NCSBN goes about re-evaluating the NCLEX-RN passing standard.

To protect the public from incompetent and unscrupulous practitioners, the government regulates certain professions by defining the scope of practice for that profession and limiting the practice of those professions to people who have adequately demonstrated that they are at least minimally competent. The governmental mechanism for restricting the practice of a profession to people who have met these requirements[1] is called licensing. Only licensees have the legal right to engage in these activities. Unlicensed practice is illegal and violators can be subjected to prosecution. In the United States, the power to issue licenses and to enforce the prohibition against unlicensed practice is typically held by the states.[2]

PSYCHOMETRIC STANDARD SETTING PROCEDURES

Given the proliferation of methods for performing standard setting exercises,[3] many people have assumed that the available methods will tell them, empirically, what the passing standard should be. This notion is mistaken. The question "how much ability is required to be competent" is not an empirical question. Because the concepts of competence and incompetence reside in individuals, the methods can only provide a procedure for coming to a reasoned consensus. However, this consensus will be specific to the particular group that made the decision. For example, if there were two different groups of experts with very different ideas of minimal competence, one would expect that the passing standards set by these two groups would be quite different. Therefore, it is important to realize that there is no passing standard that is empirically correct. Even if one used an empirical procedure to relate test scores to the probability of making a clinical error, the determination of what is an acceptable probability for error would be a personal judgment, which is likely to vary across individuals.

Nevertheless, it is possible to consider information from a variety of sources, including results from a standard setting exercise, to assist in setting a reasonable standard. For this reason, NCSBN conducts standard setting exercises that use several procedures: a modified Angoff procedure, the Beuk compromise, and two global estimates. In addition, NCSBN also collects information from other stakeholders such as educators and employers regarding the passing standard.

Who Has the Responsibility for Setting the Standard?

With regard to licensure, it is clearly the governmental body that has the authority to issue the license. For nursing in the United States, this responsibility belongs to the individual states. Often, state statutes use language that delegates this authority to a state regulatory board or to another regulatory organization. For entry-level nursing, all 60 boards of nursing have elected to use the standard set by NCSBN. Why would states choose to use a standard selected by NCSBN? The answer lies in the economies of scale[4] and the portability of test results. Because setting the standard is important, considerable resources should be expended to arrive at an informed decision. If each state or territory were individually responsible for these activities, the cost for these activities would be incurred repeatedly for very similar products. When there is a national test for jurisdictional licensing, then the cost of test development (including standard setting) is shared rather than being repeatedly

incurred by each jurisdiction. If the test reflects the same content across the nation and the same passing standard is used, then test results can be portable across jurisdictions.

Although each state board of nursing is a member of NCSBN, each state is also free to set whatever requirements it sees fit for licensing nurses. Each state is always free to use a different standard for its jurisdiction, but such an action would result in state-specific test results and the jurisdiction would be responsible for the costs of developing, implementing, and defending their state-specific standard. To facilitate the acceptance of the NCSBN standard by the states, NCSBN has purposely created a line of accountability from the NCSBN Board of Directors that sets the NCSBN standard to the states that have the power to license. More specifically, states typically have laws that establish a board of nursing and charge the members of that board with overseeing the nurse licensing process. All American boards of nursing are members of NCSBN. This membership is voluntary and each member board of nursing is permitted to send two delegates to the NCSBN Delegate Assembly to vote on issues and policies and to elect members to the NCSBN Board of Directors. The bylaws of NCSBN delegate the authority to set the NCLEX passing standard to the Board of Directors. This establishes a direct line from the state through the board of nursing through the delegates to the Board of Directors that sets the standard. To run for a position on the NCSBN Board of Directors, a candidate must either be an appointed member of a board of nursing or an employee of a board of nursing.[5] As such, each candidate is charged by his or her state with the protection of the public through the regulation of nursing. When a person accepts a position on the NCSBN Board of Directors, he or she also accepts this same charge from the other member jurisdictions.

How Does NCSBN Set the Standard?

Essentially, all licensure tests have a passing standard that must be met or surpassed before a license is awarded. Not only is it necessary to thoughtfully consider where that standard should be initially, but also it is necessary to periodically re-evaluate the appropriateness of the passing standard because practice can change over time. To ensure that the passing standard for NCLEX-RN and NCLEX-PN examinations accurately reflects the amount of nursing ability currently required to practice competently at the entry level, NCSBN's Board of Directors re-evaluates the passing standard every 3 years or when the test plan changes. In evaluating the passing standard, they consider information from a variety of sources. Although there is no limit on the information that they may consider, they are typically presented with the following information: the results of a standard setting exercise, a historical record of the passing standard and the associated summaries of candidate performance, the results from a standard setting survey, which solicits the opinions of employers and educators, and information regarding the educational readiness of high school graduates who expressed an interest in nursing.

METHOD

The NCLEX-RN Examination

The NCLEX-RN is a variable-length, computerized adaptive test. Each candidate's examination conforms to the current test plan (NCSBN, 2003) and contains 75 to 265 questions. Of these questions, 15 are unscored pretest items. Every time the examinee answers a

scoreable question, the computer re-estimates the examinee's ability and subsequently selects a question from the item bank that will both meet the test plan requirements with regard to content and have a level of difficulty that the examinee should find challenging. This provides a test that is well targeted to each examinee. After question 75 is answered, the computer attempts to determine with 95% confidence whether the examinee's true ability is above or below the passing standard. This is accomplished by determining if the candidate's ability estimate is more than 1.67 standard errors away from the passing standard. If it is above, the test stops and the examinee passes. If it is below, then the test stops and the examinee fails. If the computer cannot make a decision with 95% confidence, then it asks another question. This continues until (a) a decision is reached, (b) the maximum number of items is reached, or (c) the examinee runs out of time. If an examinee reaches the maximum number of items without a pass-fail decision being made, the 95% certainty requirement is dropped. At the maximum number of items, an examinee's ability estimate is quite precise. Ability estimates above the passing standard pass. Ability estimates at or below the passing standard fail. If an examinee runs out of time before answering the maximum number of questions, the decision process is more complex. In this case, the examinee's ability estimate on the last item is compared with the passing standard. If it is not above passing, the examinee fails. If it is above passing, then the examinee's ability estimate on the second to last item is compared to the passing standard. If this estimate is also above passing, then the third last ability estimate is compared to the passing standard. This process continues over the last 60 ability estimates. If the examinee's ability estimate drops to or below the passing standard even once on the last 60 items, the examinee fails.[6]

Every operational question in the item bank has undergone repeated review with regard to content and has met all of NCSBN's statistical requirements. The items are calibrated using a one-parameter logistic model, Rasch's (1960/1980) model for dichotomous questions.

Rasch Model

All Rasch models are logistic, latent trait models of probability for monotonically increasing functions. These models are derived not from data but from the structure necessary for measurement. Consequently, the Rasch model is imposed on data. This is quite different from "statistical" approaches in which a model is created to efficiently summarize or reproduce the observed data. The model demands that when two people of different ability encounter an item, the person with the higher ability *always* has the higher probability of answering it correctly. Similarly, when a person encounters two items of different difficulty, the more difficult item *always* has a lower probability of being answered correctly than the easier one. The philosophy behind Rasch's model is that there is a single continuum onto which both items and people are mapped. Because the items represent what the examinee can and cannot do, the ordering and relative spacing of the items articulates the construct. Subsequently, a person's ability estimate is then expressed as the point on that item continuum where the person has a 50-50 chance of correctly answering an item. It is immediately obvious that the invariance of the item hierarchy is crucial.

There are several advantages to this approach. The first advantage is that aberrant responses can be identified. When a person's responses to the items or the responses to an item by the population deviate noticeably from the expectations of the model, then that person or item should be examined. Statistical models that use additional parameters to model these deviations conceal the underlying problem and destabilize the construct. Wright (1997a, 1999) provides a clear explanation of how these extra parameters (discrimination

and pseudo-guessing) destabilize the construct across the ability continuum to the extent that they make a difference. The typical rationale for using these multiparameter models is that they better represent the data, which they indeed do, but at the expense of detecting anomalous responses.

A second advantage is that because all items and people are located on the same continuum, when the data fit the model, the probability of a correct response to every item on the continuum is known, once the person's ability is established. This includes items that the person was never asked. A third advantage is that changes in the population of examinees, either improvement or deterioration, can be measured so long as construct, as defined by the items, maintains their same relative positions as a whole. The construct remains stable so long as the relative difficulty among the items remains unchanged. Conversely, changes in the construct over time can be detected by changes in the item hierarchy. There are many more advantages (Wright, 1997a, 1997b, 1999) that have not been addressed here.

The dichotomous Rasch model specifies that the probability of a correct response is governed by the difference between the ability of the person, β_v and the difficulty of the item, δ_i. However, the difference $(\beta_v - \delta_i)$ can range from infinity to negative infinity, but the probability of a correct response is limited to the range of 0 to 1. Converting the probability to a log odds ratio solves the restriction of range problem. Expressed mathematically, the dichotomous Rasch model is specified as:

$$\Pr\{(x_{ni1})|\beta_n, \delta_i\} = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}} \tag{1}$$

where $\Pr_{ni1}$ is the probability of a correct response ($\Pr_{ni0}$ would be the probability of an incorrect response), β_n is the ability of person n, δ_i is the difficulty of item i, and e is the base of the natural log function.

Because the model requires that the relative difficulty of the items remains stable, responses by individuals or groups of individuals that grossly violate that notion can be detected statistically through a variety of procedures such as person misfit, item misfit, parameter drift, differential item functioning, and others.

Rasch's model separates the person and item parameters, yet expresses them on the same scale. As a result, the same person ability estimate should be derived regardless of the particular items administered. This is true regardless of the overall difficulty of the test. Similarly, item difficulty calibrations should be the same regardless of the particular people who answered the question. This is true even when items are calibrated on groups of people with noticeably different mean abilities. Notice that the requirements of sampling theory (random assignment to create equal groups, normal distributions, interval scale observations, etc.) are not required for the Rasch model. When the responses fit the Rasch model, interval measurement is achieved and a stable construct is articulated for the entire functional range of items.

What Types of Information Should Be Considered?

What types of information should the Board of Directors consider when setting the passing standard? The board may look at many types of information, but it has a responsibility to weigh and interpret that information with regard to protecting the public. When re-evaluating the passing standard, the NCSBN Board of Directors considered a variety of factors. Typically this includes reviewing: (a) background information related to what the

standard has been historically (history of the passing standard, associated pass rates, and indicators of academic readiness), (b) information regarding the opinions of educators and employers with respect to the competence of recent graduates and licensees, and (c) information regarding the opinions of content experts with respect to how well minimally competent examinees should perform on a prespecified, nonadaptive form of the test.[7] To facilitate passing standard comparisons, all passing standards (current, previous, and suggested) are expressed on the NCLEX-RN ability scale.

Using the Rasch Model as a Framework for Comparisons

The Board of Directors considers a variety of perspectives regarding where the passing standard should be. Some of the information merely indicates whether a particular group thought the standard should be raised, lowered, or retained. Other types of information are more specific and can be transformed into a suggested passing standard. NCSBN uses the Rasch model to calibrate every question on the NCLEX and to generate person ability estimates for every examinee. Consequently, the passing standard must also be expressed in terms of this continuum.

To accomplish this, NCSBN uses the Rasch model to provide a unified framework, the ability continuum, upon which each suggestion is placed. In this way, the Board of Directors can consider and weigh a variety of perspectives. Specifically, the raw score results from a standard-setting exercise can be translated to values on the ability continuum. Future pass rates can be predicted by imposing these values as cut-scores upon a representative ability distribution. Conversely, designated pass rates (as suggested by the employer and educator survey) can be imposed upon recent data sets to find the value that will produce that pass rate. Also, historical pass points can be included as well.

Background Information

A chart that illustrates the pass rate by quarter since April 1994 was created. This chart also indicates the particular passing standard that was in effect at any given time. Also, a table containing the mean ACT Assessment® composite score by graduation year was created.

Examinee-Based Opinions

Every year, nurse educators and employers of newly licensed registered nurses are surveyed regarding their opinions about the competence of the current cohort that is entering the workforce. Four different types of nursing professionals are surveyed: administrators of nursing education programs, directors of nursing at hospitals, long-term care facilities, and community/home health agencies.

Content-Based Opinions

In September 2003, a nine-member standard setting panel was convened to review a 300-item test form with regard to how well a candidate should perform on this form to be considered at least minimally competent. All nine raters were licensed registered nurses and represented the four National Council geographic areas. Two of the raters were members of an ethnic minority group and one of the raters was a newly licensed registered nurse.

A modified-Angoff (Angoff, 1984) technique was used in the workshop to develop a proposed passing standard. This technique requires the panel members to estimate the performance of minimally competent candidates on individual items. Specifically, the panel members were asked, "Out of 100 minimally competent candidates, how many of them

would answer this item correctly?" Because the rating process is difficult for raters to grasp without detailed instruction, the raters completed a thorough training process before beginning the actual rating of items. During this training stage, the raters discussed the defining characteristics of a "minimally competent entry-level RN." The final part of the training workshop focused on a discussion of item difficulty. The panel members developed their ability to judge items by going through the actual rating procedure for a set of 30 total practice items that were representative of the new NCLEX-RN test plan. Feedback and discussion was provided in order to enhance the rating process.

After the training, the raters estimated the performance of minimally competent candidates on a reference form of 300 real items. This reference form was designed to meet the requirements of the new test plan and to approximate the difficulty distribution found in the current operational item pool. Raters evaluated items in sets of 25. After assigning a rating, the ratings were then compiled, summarized, and displayed for the group. These summaries were then used as a basis for discussion among panel members. The discussion included reasons for why the discrepancies in the ratings may exist among members. After discussing each item, the raters were at liberty to change their rating. Once a set was completed, the next set of items was rated, summarized, and discussed. This process continued until all 300 items on the reference form were rated. Throughout this process, raters were allowed to compare their predictions of minimally competent candidate performance with statistical predictions of expected candidate performance for candidates who are near the current passing standard. The "modified-Angoff" passing standard recommendation is simply the mean of the group's item judgments.

After the Angoff procedure was completed, the raters were asked to respond individually to three questions. In an effort to tap into their expertise at a global level, and to solicit opinions that were not influenced by any pressure for consensus, judges were asked:

1. "What percentage of reference group (first-time, U.S.-educated) candidates do you think presently fail the NCLEX-RN?"
2. "What percentage of reference group candidates do you think are not competent to practice?" and
3. "Of the 300 items you have just reviewed, what percentage of those items do you think a candidate needs to answer correctly in order to demonstrate minimum competence?"

The Beuk (1984) compromise is a procedure that bridges the gap between content-based judgments about competence (the modified Angoff recommendations, the judges' perceptions of how minimally competent candidates would perform on the reference form) and judgments about the competency of the population taking the NCLEX-RN. The Beuk calculation uses the results from both the modified-Angoff and Global Question #2 to approximate the judges' perception of how minimally competent candidates would perform.

Assessing Impact

To predict the impact of imposing different passing standards, two cumulative frequency distributions were created. Using data from RN candidates who tested from August 1, 2002, through July 31, 2003, a distribution composed of first-time, U.S.-educated candidates and a distribution for all candidates were created. Using these distributions, any suggested passing standard can be evaluated with regard to an expected pass rate. Although there is no specified pass rate that is being sought, this piece of information can prevent the adoption of an unrealistic or unreasonable passing standard. These predictions about the pass rates associated with different passing standards make a few assumptions. They are as follows:

(a) the ability of the candidate pool, both the reference group and total population, will be the same as it has been in the last year. Although schools and even regions may see some variability from year to year, when considering all the member boards of nursing together there is very little change from year to year. (b) The most recent data (for a full year) will be the most reflective of future annual ability distributions. (c) The candidate ability estimate that is based upon the entire test will provide a better estimate of the candidate's ability than will a shortened version. (d) Some examinees who run out of time on the test will fail via the "last 60 rule" despite having a final theta that is above passing. Because the pass rate is based upon an expected theta distribution and some thetas that are above passing will fail, there will be a slight bias in the predicted pass rate. Pass rates for the reference group and total population will likely be lower by 0.35% and 0.65% respectively.

RESULTS

The actual results were presented to the NCSBN Board of Directors for consideration. After considering the results, the Board of Directors deliberated and selected a revised passing standard. However, the author does not have the authorization to release the actual results. *Therefore to illustrate the process, the author has fabricated results that are fairly similar to the actual results*.

Background Information

To provide the Board of Directors with a background regarding the current state of affairs, the following information was presented. This information is provided by year (Figure 1), by quarter for (reference group) first-time, U.S.-educated examinees (Figure 2) and by quarter for all examinees (Figure 3).

   Also, provided is information on the academic readiness of high school graduates, mean ACT Assessment scores, (Table 1) of those who have expressed an interest in nursing. This information can be helpful with regard to explaining unexpected changes in pass rates. It should be noted that the Board of Directors uses this information only in an explanatory way and never as an indicator of the meaning of minimal competence. More specifically, the Board does not use increases or decreases in the ability of these high school students as a contributing factor to increase or decrease the passing standard.

Examinee-Based Opinions

Every year, a sample of nursing professionals is sent a written survey that solicits their opinions concerning the competence of the current cohort of entry-level registered nurses. The 2000, 2001, and 2002 surveys were mailed to four different samples of nursing professionals: administrators of nursing education programs, directors of nursing at hospitals, long-term care facilities, and community/home health agencies. Each year, a total of 1,600 RN surveys were sent; 400 to each group. Returned surveys that indicated that they did not have contact with entry-level nurses were excluded from further analysis. Table 2 lists the number of surveys returned and analyzed.

   In one question, the respondents were given four categories (a 2×2 table, competence by pass-fail status) and asked how many of the new RNs that they had contact with fell into each category (Table 3). Educators reported that they felt 92.0% of their graduates were competent, regardless of their pass-fail status on the NCLEX-RN. To produce a pass rate among first-time, U.S.-educated candidates that would match this expectation, the
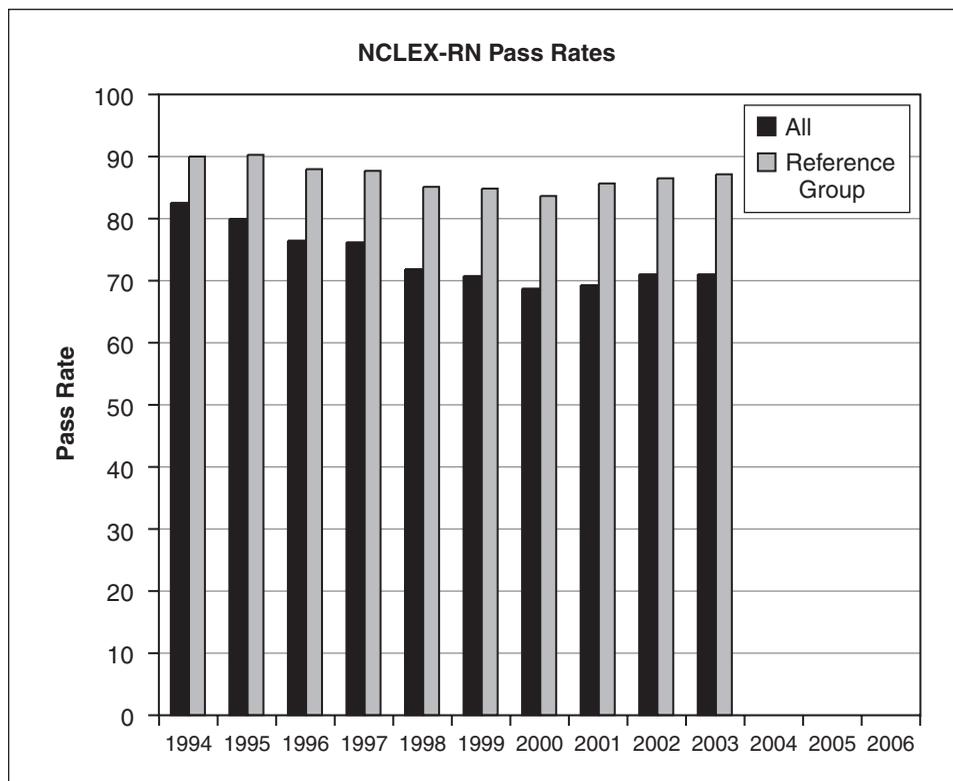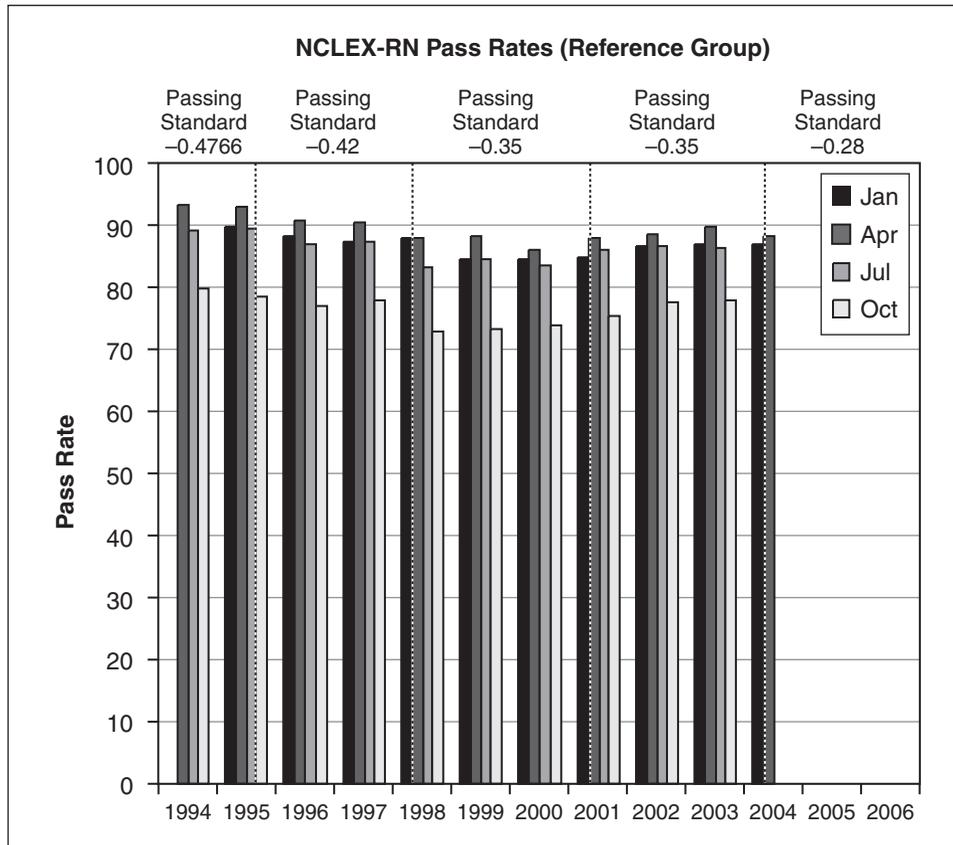
Figure 1. Annual pass rates on the NCLEX-RN: Reference group and total candidate population.

passing standard would have to be set to approximately -0.49 logits. This standard would be noticeably less stringent than the existing passing standard (-0.35). The employers felt that 93.0% of the newly licensed nurses were competent. Because only candidates that pass the examination are typically licensed, responses for the failing candidates cannot be interpreted. To produce a pass rate among all passing candidates that would match this expectation, the passing standard would have to be set to approximately -0.26 logits, which is slightly more difficult than the present standard. Note that because the employers can only evaluate people who have already met the existing standard, any percentage less than 100% is a recommendation for a more stringent standard.

The survey also asked each respondent his or her opinion regarding the quality of the current RN cohort as compared to previous cohorts. In this regard, more than half of the respondents tended to consider quality of the cohort as remaining the same (Table 4). This was true for both educators and employers. Similarly, when asked their opinions regarding the passing standard, the majority of respondents felt that the passing standard should remain the same (Table 5). However, there did seem to be a stronger tendency among employers to want to raise the standard.

Examinee-based judgments were also solicited from members of the standard setting panel. When the raters were asked "what percentage of first-time, U.S.-educated candidates were not competent to practice?" (Table 6, global estimate—reference group), the average percentage was 21.1%. This implies that the panel thought that only 78.9% of the

Figure 2. Quarterly pass rates on the NCLEX-RN: First-time, U.S.-educated (reference group).

first-time, U.S.-educated candidates should have passed the NCLEX. Although this information is one of the factors used in the Beuk Compromise, this global estimate can also be directly transformed onto the NCLEX-RN scale. Assuming the August 2002 through July 2003 ability distribution for first-time, U.S.-educated candidates is typical, a pass rate of 78.9% corresponds to a theta of -0.16.

Content-Based Opinions

The results of the Angoff procedure were not as homogenous as expected. Regarding the percentage of items that a minimally competent candidate would answer correctly on the reference form, the raters' judgments ranged from 53% to 69%. The average was 58%, which translates to -0.01 logits on the NCLEX-RN scale. A candidate whose ability was at the current passing standard (-0.35 logits) would have answered only 50.1% correctly. The -0.01 passing standard is expected to generate a 67.0% pass rate for first-time, U.S.-educated candidates and a 52.8% pass rate for the total population. However, two of the raters (raters 6 and 7) were noticeably more severe than the other seven. Excluding those two raters, the average was 55.3%, which translates to -0.14 logits on the NCLEX-RN scale. This passing standard is expected to generate a 78.6% pass rate
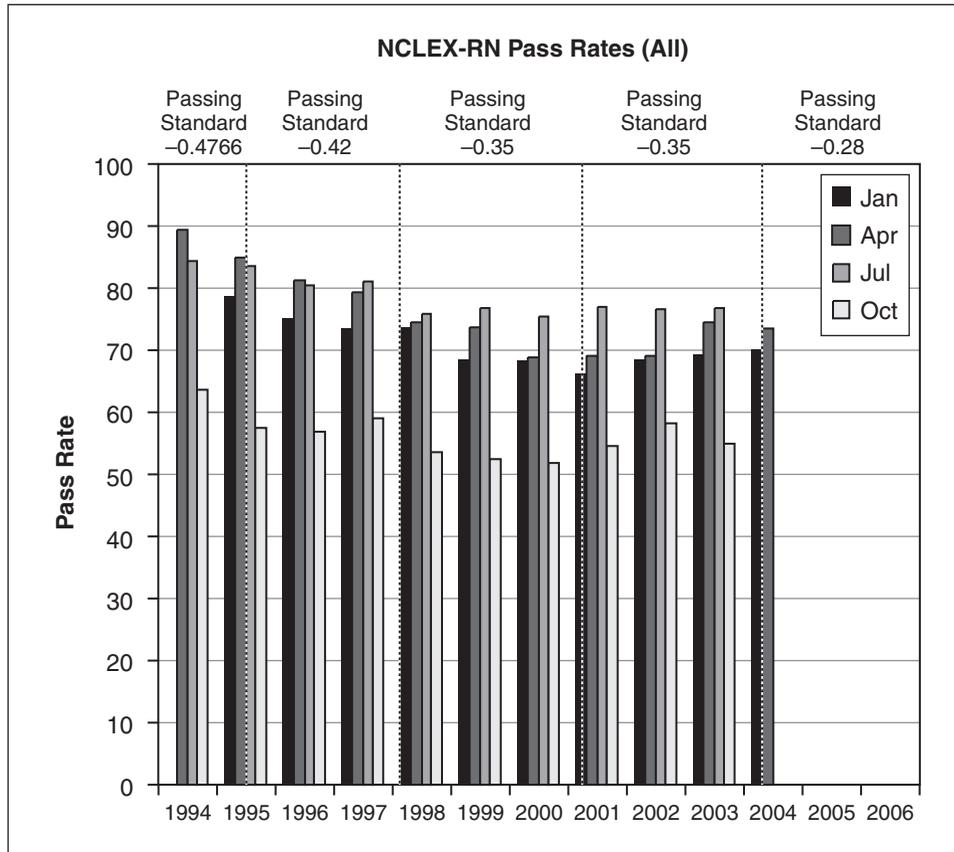
Figure 3. Quarterly pass rates on the NCLEX-RN: Total candidate population.

TABLE 1. ACT Composite Scores of High School Graduates Who Expressed an Interest in Registered Nursing

High School Graduation Year	Mean ACT Assessment® Composite Score	N
1990	18.5	1,228[a]
1991	18.6	1,361[a]
1992	18.6	1,675[a]
1993	18.8	1,962[a]
1994	18.7	2,013[a]
1995	18.6	1,968[a]
1996	18.5	1,799[a]
1997	18.6	17,938[b]
1998	18.6	16,707[b]
1999	18.6	16,458[b]
2000	18.6	16,651[b]
2001	18.6	16,822[b]
2002	18.5	19,116[b]
2003	18.5	25,512[b]

[a]N = 10% sample of all students who tested on the national date and listed a valid HS code and an interest in registered nursing. [b]Total sample of all students who tested on the national date and listed a valid HS code and an interest in registered nursing.

TABLE 2. Response Rate for Survey of Educators and Employers of Registered Nurses

	Surveys Returned				Surveys Analyzed			
	2000	2001	2002	Total	2000	2001	2002	Total
Employers								
Community/home health care	64/400	54/400	38/400	156/1200	4/400	9/400	2/400	15/1200
Long-term care	57/400	39/400	41/400	137/1200	29/400	10/400	24/400	63/1200
Hospital	56/400	35/400	47/400	138/1200	42/400	23/400	43/400	108/1200
Employers total	177/1200	128/1200	126/1200	431/3600	75/1200	42/1200	69/1200	186/3600
Educators	74/400	51/400	121/400	221/1200	42/400	28/400	100/400	170/1200

TABLE 3. Illustrative Survey Data: Counts of RNs by Pass-Fail Status and Perceived Competence[a]

	Employers		Educators	
	Competent	Noncompetent	Competent	Noncompetent
Pass	980/1,040 (93.0%)	60	5,177	69
Fail	—	—	365	413
Total	—	—	5,542/6,024 (92.0%)	482

[a]Because employers should be evaluating only licensed RNs, the percentages for failers cannot be interpreted; therefore the competency percentage is based on only passers. Educators evaluate all their graduates; therefore it is appropriate to include both passers and failers in the cohort.

TABLE 4. Illustrative Survey Data: How Does the Current RN Cohort Compare to Previous RN Cohorts?

	Worse	Same	Better
Educators	28%	51%	21%
Employers	23%	60%	16%

TABLE 5. Illustrative Survey Data

The NCLEX-RN Passing Standard Should:

	Be Lowered	Stay the Same	Be Raised
Educators	8%	82%	10%
Employers	1%	75%	24%

for first-time, U.S.-educated candidates and a 63.1% pass rate for the total population. The reliability between the raters on their final ratings was adequate (0.89).

The results from the global assessments were also less homogenous than expected. The results of these three questions were as follows: (a) The raters' estimates of the percentage of candidates currently failing the NCLEX-RN examination ranged from a low of 10% to a high of 30%, with an average of 19.9%. These values are close to what is actually observed. Typically, 15% of the reference group and 30% of the total population fails. (b) Their estimates of the percentage of candidates who actually are not competent, and who should therefore fail the examination, ranged from 5% to 50%, with an average of 21.1%. Using the August 2002 through July 2003 reference-group ability distribution, failing 21.1% of the group occurs at -0.16 logits on the NCLEX-RN scale. (c) The raters' global estimate of the necessary percent correct (of the 300 items on the reference form) to demonstrate minimum competence averaged 59.1%, with a range from 50% to 80%. This translates to 0.30 on the NCLEX-RN scale.

The judges' answers to the question, "What percentage of the first-time, U.S.-educated population are actually not competent?" were used as the basis for the Beuk compromise procedure. A compromise rating resulting from this procedure produced a raw score of

TABLE 6. Projected NCLEX-RN Pass Rates Based on Different Passing Standards

		Expected Annual Pass Rate	
	Passing Standard	NCSBN Reference Group[a]	NCSBN Total Group[b]
Current standard	-0.35	87.4%	72.6%
Revised standard	-0.28	84.0%	68.7%
Panel of judges[c]			
Modified-Angoff (all 9 raters)	-0.01	67.0%	52.8%
Rater 1 (160 of 300 items, 53%)	-0.21	80.6%	65.0%
Rater 2 (162 of 300 items, 54%)	-0.19	79.8%	64.2%
Rater 3 (166 of 300 items, 55%)	-0.14	78.6%	63.1%
Rater 4 (176 of 300 items, 59%)	0.00	65.9%	51.8%
Rater 5 (175 of 300 items, 58%)	-0.02	68.0%	53.7%
Rater 6 (207 of 300 items, 69%)	0.45	19.7%	14.7%
Rater 7 (206 of 300 items, 69%)	0.42	21.6%	16.1%
Rater 8 (159 of 300 items, 53%)	-0.23	81.5%	66.0%
Rater 9 (164 of 300 items, 55%)	-0.16	78.9%	63.4%
Angoff (excluding 2 severe raters)	-0.14	78.6%	63.1%
Beuk compromise	-0.13	78.2%	62.7%
Global estimate (items)	0.30	30.4%	22.9%
Global estimate (reference group)	-0.16	78.9%	63.4%
Survey of professionals[d]			
Employers	-0.26	83.0%	67.6%
Educators	-0.49	92.0%	78.7%

[a]The NCSBN Reference Group is based on the ability distribution of first-time U.S.-educated examinees that tested from August 1, 2002, through July 31, 2003. (This represents an entire year of examinations.) The pass rate is based on the examinee's final ability estimate. However, because there are instances in which candidates with a final theta above passing fail via the "last 60 rule," this estimate is likely optimistic. A downward adjustment of 0.35% (272/78,716) may be warranted. [b]The NCSBN Total Group is based on the ability distribution of all examinees that tested from August 1, 2002, through July 31, 2003. (This represents an entire year of examinations.) The pass rate is based on the examinee's final ability estimate. However, because there are instances in which candidates with a final theta above passing fail via the "last 60 rule," this estimate is likely optimistic. A downward adjustment of 0.65% (806/124,359) may be warranted. [c]All results from the Panel of Judges have been fabricated, but are similar enough to the actual results to provide a useful illustration of the process. [d]The Survey of Professionals results have been fabricated, but are similar enough to the actual results to provide a useful illustration of the process.

166 on the 300-item reference form, translating to a -0.13 ability estimate that would result in a 78.2% pass rate for first-time, U.S.-educated candidates and a 62.7% pass rate for the total population.

Comparing the Standards and Assessing the Impact

When there is a common scale such as the Rasch-derived NCLEX-RN scale and an expected ability distribution, it is possible to predict the pass rate that any particular passing standard will produce. Of course, this is limited by the stability of the scale and the accuracy

with which the expected ability distribution (based on recent trends) matches the ability distribution found in the near future. In this case, the NCLEX-RN is a very stable scale and the ability distribution of all candidates and of first-time, U.S.-educated candidates doesn't vary drastically from year to year. Therefore, it is reasonable to assume that the ability distribution observed from August 1, 2002, through July 31, 2003, should closely match the ability distribution that will be observed during the 12 months following the implementation of a new standard (April 1, 2004, through March 31, 2005). Table 6 and Figure 4 summarize the different suggested passing standards and the predicted pass rates associated with each one.

Board Deliberation

The NCSBN Board of Directors met in December 2003 to review the NCLEX-RN passing standard. Using this information and their own personal experience as regulators, the board decided to raise the standard by 0.07 logits from -0.35 logits to -0.28 logits. The guiding principle in this discussion was how high must the standard be to protect the public. The board concluded that the standard should be raised in response to changes in U.S. health care delivery and nursing practice that have resulted in the increased acuity of clients seen by entry-level RNs.
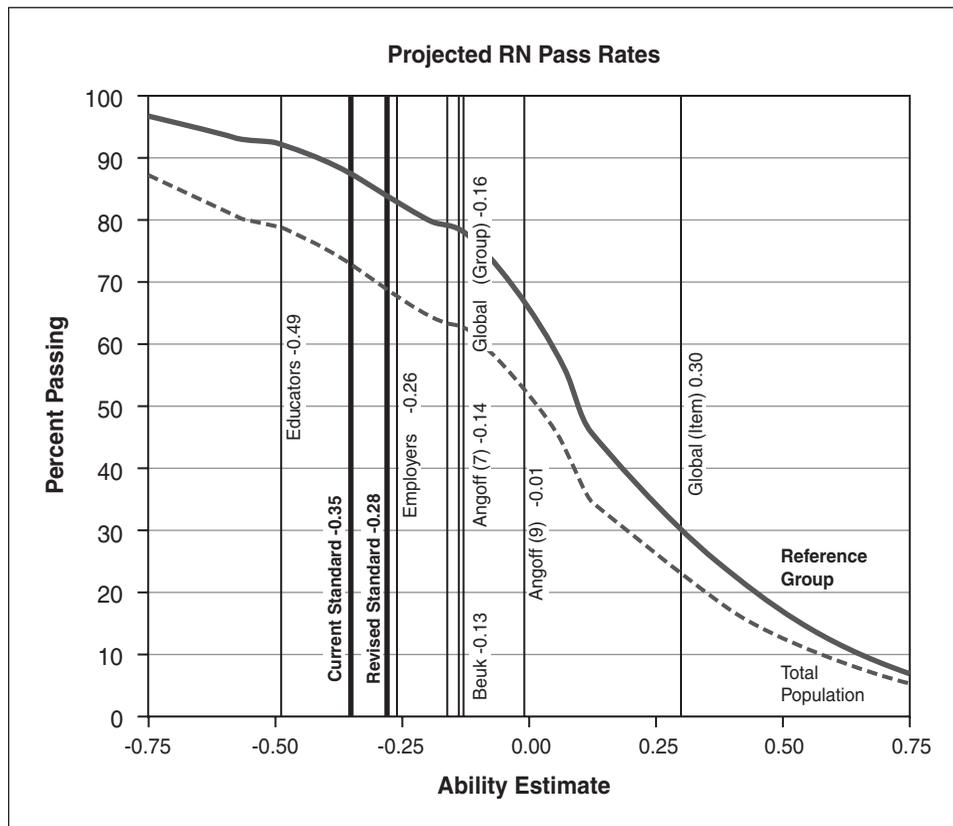
Figure 4. Projected annual RN pass rates (reference group and total population) based on different (fabricated) passing standards.

This new standard was implemented April 1, 2004. Using the candidate ability distribution from August 1, 2002, through July 31, 2003, as the expected ability distribution for April 1, 2004, through March 31, 2005, it is anticipated that approximately 68.7% of the total candidate population and 84.0% of the first-time, U.S.-educated candidates will pass. This is a drop of about 2.5% to 3.0%. Based upon one quarter's data, the pass rate is lower but by less than the expected amount. The pass rate for first-time, U.S.-educated candidates in April–June 2004 (88.5%) was down by 1.6% from April–June 2003 (90.1%). The pass rate for the total candidate population in April–June 2004 (73.4%) was down by 0.9% from April–June 2003 (74.3%).

DISCUSSION

One advantage of publishing a standard setting study is that it publicly demonstrates that the standard was set after giving consideration to many relevant perspectives. Having this type of study available and a brief written statement by the group setting the passing standard can go a long way toward demonstrating that the standard was not set capriciously. NCSBN's procedures are a quality example of how to set passing standards for licensing examinations. For this reason, it is important to share them in an academic format so that these ideas can be adapted to the needs of other organizations and enhancements can be suggested. The types of data considered can be classified into three categories: historical performance, directional recommendation, and a specific recommendation.

Historical

When a test is brand new, an historical review of the passing standard isn't possible, but for more established tests, an historical review of where the standard has been can provide an experiential context in which the old standard can be evaluated. After 3 years of using the existing standard, do we think the standard needs to be raised or lowered? This is why the history of both the passing standard and candidate performance on the examination since 1994 is provided. This information should be considered in conjunction with the changing requirements of the profession. Because the nature of the work may change, the definition of minimum competence may change as well. For example, suppose the combined effects of a nursing shortage and managed care's cost reduction strategies have caused the typical work environment to be characterized by less supervision of new nurses by senior nurses. Further suppose that there is an increase in the number of patients and in the seriousness of their illnesses. In this case, the change in the work environment might warrant raising the passing standard. On the other hand, if the legal scope of practice suddenly became more limited or some complex procedures became automated and less prone to error, then this might warrant lowering the standard.

Directional Recommendation

Some of the survey questions asked the respondent (an educator or employer) whether he or she thought the standard should be raised or lowered, but without generating a specific recommended passing standard. At first glance, these types of questions appear to be less powerful than those that generate a specific recommended passing standard. However, they are helpful in interpreting the recommended passing standards. For example, employers

thought that 93.0% of their newly licensed nurses were competent. Passing only 93.0% of the population that had passed under the present standard would require a passing standard of -0.26 logits, which is slightly more difficult than the current standard of -0.35 logits. Please note that because we are considering only those candidates who passed, any percentage less than 100% must result in a more difficult standard. When asked directly whether the standard should be raised, lowered, or retained, 75% said that it should be retained and 24% said that it should be raised. Both approaches produced congruent results. Yet this is not always the case. For instance, on average educators thought that 92.0% of their students were competent. To pass 92.0% of first-time, U.S.-educated candidates, the passing standard would have to be -0.49 logits, noticeably easier than the current standard. However, 82% of the educators said that they thought the current standard should be retained. The reason for such a discrepancy is up for interpretation. Perhaps some educators consider labeling their students as incompetent to be much too harsh. Perhaps admitting that a low percentage of their students are competent is a poor reflection on their program. Certainly there are other potential explanations as well. Nevertheless, examining the results from both procedures alerts the Board of Directors to the discrepancy.

Integrity of the Board of Directors

The integrity of any licensing or certification organization's Board of Directors is very important. The board should not abdicate the responsibility for making the decision to another committee because the Board of Directors is legally accountable for the placement of the standard. The Board of Directors should be making this decision keeping in mind the public's safety. In this regard, apparent conflicts of interest are problematic. For example, if the people who set the passing standard are directly accountable to an organization whose purpose is to promote the well-being of the profession, then the Board of Directors may appear to be setting a standard that protects the profession rather than the public. Even when the Board of Directors is doing its utmost to set the standard to protect the public, it is still vulnerable to that accusation. NCSBN is an organization that is not accountable to nursing societies and associations, making it resistant to those types of allegations.

Conclusion

Different groups can have different perspectives regarding how stringent the passing standard should be. Some educators may feel that a high percentage of their students should pass the test on their first attempt and that the passing standard should reflect that perception. Some employers may feel that the standard is too low because the newly licensed nurses that they encounter are "too green." Other employers may feel that the standard is too high because there doesn't seem to be enough nurses in the labor market to fill their open positions. Collective bargaining units may feel that a higher standard would be beneficial because it would reduce the supply of available nurses and therefore afford them a better bargaining position. These are all valid concerns, but each represents the interests of different parties. However, state and territorial boards of nursing and NCSBN are charged with protecting the public's safety by regulating the practice of nursing. For this reason, NCSBN considers the perspectives of educators and employers regarding the passing standard, but gives the responsibility for setting the national standard to the NCSBN Board of Directors.

NOTES

1. In addition to minimum competency requirements (e.g., age requirements, education/training requirements, passing a minimum competency test, etc.), licensing programs usually have other components that preclude or remove people from practice for unethical or unprofessional behavior (e.g., criminal background checks, disciplinary procedures, etc.).

2. However, the federal government licenses some professions because they are inherently related to interstate commerce (e.g., commercial pilots, air traffic controllers, etc.).

3. There are many different procedures that can be used for standard setting exercises. Livingston and Zieky (1982) describe several popular methods such as Nedelsky's method, Angoff's method, Ebel's method, borderline group method, contrasting groups method, and the up-and-down method. Cizek (2001) provides a more comprehensive review of standard setting methods.

4. "Economies of scale" refers to the phenomenon that often occurs in mass production. Because the fixed costs of a test such as setting a passing standard or paying a printer to typeset an exam booklet are the same for one exam administration or 1 million exam administrations, the cost per exam drops as the number of exams administered rises.

5. Only jurisdictions that are members of NCSBN can send delegates or run for office.

6. An ability estimate at any point on the test is based upon the responses to all items up to that point. Therefore, it would be incorrect to say that the "last 60 rule" considers only the responses on the last 60 items. It is also important to keep in mind that the maximum item rule and the "last 60 rule" are essentially a second chance for those examinees that were not able to meet the requirement of demonstrating their competence with 95% certainty.

7. Other types of information can also be considered, such as the educational readiness of high school students who indicate an interest in nursing on their college entrance examination. This last type of information doesn't really express an opinion regarding where the standard should be placed, but it could help to explain an unexpected change in passing rates.

REFERENCES

Angoff, W. H. (1984). *Scales, norms and equivalent scores*. Princeton, NJ: Educational Testing Service.

Beuk, C. H. (1984). A method for reaching a compromise between absolute and relative standards in examinations. *Journal of Educational Measurement, 21*, 147–152.

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

Livingston, S. A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment test*. Chicago: MESA Press. (Original work published 1960).

Wright, B. D. (1997a). A history of social science measurement. *Educational Measurement: Issues and Practice, 16*(4), 33–45, 52.

Wright, B. D. (1997b). Fundamental measurement for outcome evaluation. *Physical Medicine and Rehabilitation: State of the Art Reviews, 11*(2), 261–288.

Wright, B. D. (1999). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement* (pp. 65–104). Mahwah, NJ: Lawrence Erlbaum.

acknowledge the hard work of the panelists who provided thoughtful responses throughout the process. Finally, I would like to acknowledge the Board of Directors, which has shouldered the responsibility for considering all this information and deciding upon a passing standard.

Author Guidelines for Journal of Nursing Measurement

The Journal of Nursing Measurement invites manuscripts directed to the interest of nurse clinicians, educators, and researchers. The purpose of the Journal is to serve as a forum for the dissemination of information regarding instruments, tools, approaches, or procedures developed or utilized for the measurement of variables for nursing practice, education, and research. Particularly welcome are manuscripts of studies on the development and testing of nursing measures, and evidence for the reliability and validity or sensitivity and specificity of such instruments. Innovative discussions of theories, principles, practices, and issues relevant to nursing measurement are invited as well. Letters and comments on published articles are welcome.

Manuscripts will be acknowledged on receipt. *Journal of Nursing Measurement* is a refereed journal. Following preliminary review by the Editors, each manuscript will undergo blind peer review by two experts in the methodology and the topic area of the manuscript. Articles are reviewed by members of the editorial review board comprised of measurement experts and also by individuals from the panel of manuscript reviewers who have backgrounds in a variety of clinical and functional areas. On publication, each author will receive a copy of the Journal in which his/her manuscript appears.

The following are guidelines for developing and submitting a manuscript. Manuscripts that do not conform to these guidelines will be returned to the author without review.

Manuscripts and Other Requirements

1. Authors should submit their manuscript as an email attachment in rich text format (RTF) or MS Word 2000 to: journalnm@nursing.emory.edu. Mail submissions should include one original and three copies of each manuscript, along with one computer disk in rich text format (RTF) containing the full manuscript and supporting files (i.e., tables, instruments, pictures, etc.). A copy of the instrument(s) may be included in an appendix. Instruments will be published at the discretion of the editors.
2. Manuscripts should be professionally prepared in accordance with the *Publication Manual of the American Psychological Association*, 5th Edition, 2001.
3. An abstract of approximately 125 words should include the following headings: (a) Background and Purpose, (b) Methods, (c) Results, (d) Conclusions, and (e) 4 to 6 keywords.
4. Letters to the Editor should be under 300 words. Those accepted for publication may be edited or abridged.
5. The original document should be size 8.5" x 11".
6. Double-space everything, including references, quotations, tables, and figures.
7. Leave generous margins (at least one inch all around) on each page.
8. Type should not exceed 18 characters per inch.
9. Avoid footnotes whenever possible.
10. Quotations of 300 words or more from one source require written permission from the copyright holder for reproduction. Adaptation of tables and figures also requires reproduction approval from the copyrighted source. It is the author's responsibility to secure such permission, and a copy of the publisher's written permission must be provided to the Journal editors immediately upon acceptance of the manuscript for publication.
11. Place authors' names, positions, titles, place of employment, mailing addresses, and email addresses on the cover page so that the manuscripts may be reviewed anonymously.
12. All figures must be submitted in camera-ready form. In addition, photos and line art figures should be sent on disk (TIFF at 300 ppi or EPS at 800 ppi).
13. Submit manuscripts to: Ora L. Strickland, PhD, RN, FAAN, Senior Editor, *The Journal of Nursing Measurement*, Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Road, NE, Room 366, Atlanta, GA 30322. Phone: (404) 727-7941; Fax: (404) 727-0536 or (404) 712-8872. E-mail: journalnm@nursing.emory.edu

Note: Authors bear full responsibility for the accuracy of references, quotations, tables, and figures. Upon acceptance of the article authors are expected to fill out the copyright agreement form and mail it to the editor.

Checklist for Submission

_____ Cover letter

_____ Computer disk or email attachment in Rich Text Format (RTF). Disk must contain the full manuscript in one file; tables in another file; and other supporting files. Files should be prepared in PC-compatible word processing software, preferably MS Word.

_____ Acknowledgments

_____ Permission letter(s) (for reproduction of copyrighted materials only)

_____ Manuscript: original and three copies (if sent by mail)

_____ Author identification page with email address

_____ Title page

_____ Abstract (125 words) includes background and purpose, methods, results, conclusions, and keywords

_____ Introduction; includes purpose and scope of manuscript

_____ Text pages

_____ References

_____ Figures or tables with legends

Manuscript Requirements

Cover page:
Place manuscript title, and authors' names, positions, titles, place of employment, mailing addresses, and email addresses on the cover page so that the manuscripts may be reviewed anonymously.

Title page in this order:
1. A short running head of no more than 40 characters (count letters and spaces)
2. Title of the article, which should be concise but informative
3. Date of submittal (include date of revision submittal if applicable and *JNM* manuscript number assigned).
4. Abstract

Abstract:
The abstract should appear on the title page and should be no more than 125 words. The abstract should briefly state the background and purpose of the study or investigation, basic methods and procedures (selection of study subjects, analytical methods), main findings (summarize specific reliability and validity data), and the principal conclusions. It should emphasize new and important measurement aspects of the study or observations. The abstract will include the following headings: (a) Background and Purpose, (b) Methods, (c) Results, (d) Conclusions, and (e) Keywords.

Keywords:
Below the abstract, authors should provide 4 to 6 keywords or short phrases that will assist indexers in cross-indexing the article. Keywords are published with the article. Terms from the medical subject headings (MeSH) list of *Index Medicus* should be used, if at all possible.

Manuscript Content Requirements:
Data-based measurement manuscripts should include the following headings: (a) Introduction (purpose and scope of manuscript), (b) Background and Conceptual Framework, (c) Procedures for Instrument Development, (d) Description, Administration, and Scoring of the Instrument, (e) Methods (sample, procedures, and approaches to reliability and validity assessments), (f) Results (findings from reliability and validity assessments), (g) Discussion (measurement-focused interpretations of findings, limitations, and conclusions), (h) References, and (i) Figures or tables with legends.