



October 2011

Understanding the NCLEX® Examination Through the Core Values of NCSBN

Introduction

The National Council of State Boards of Nursing, Inc. (NCSBN®) is a not-for-profit organization that is composed of 60 jurisdictional boards of nursing (BONs) in the U.S. and its territories. NCSBN was established to provide an organization through which BONs act and counsel together on matters of common interest and concern affecting public health, safety and welfare, including the development of licensing examinations in nursing (O'Neill, Marks, & Reynolds, 2005). To assist BONs in making licensure decisions and as a critical element in ensuring protection of the public, 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®). The development, administration and future direction of these examinations is guided by the core values of NCSBN, which include collaboration, excellence, innovation, integrity and transparency in advancing regulatory excellence worldwide¹. This article is intended to provide information about the development and administration of the NCLEX® examination as it relates to the core values of NCSBN.

Licensure, Certification and Education Examinations

In order to understand the rigorous development and structured administration processes associated with the NCLEX examinations, it is important to distinguish the differences between a high stakes licensure examination versus an examination developed for certification or educational purposes. Licensure is the act of granting a legal right to a qualified individual to practice in a particular job or profession (Impara, 1995; Schoon & Smith, 2000). The purpose of licensure is to protect the public from mental, physical or emotional harm by practitioners who may not be sufficiently competent to enter the profession for which the license is granted (Impara, 1995). Licensure has the inherent property of instilling trust within the public that the licensee is competent and has met the initial requirements for entry into practice as specified by a regulatory entity. Typically, to be licensed into a profession, the practitioner must successfully complete an approved educational program and an examination. Because the licensure examination is one of the major hurdles for entry into practice, it is considered to be a high-stakes examination and therefore necessitates high standards to ensure the examination is valid (i.e., measuring what it purports to measure).

Certification is defined by the U.S. Department of Health, Education, and Welfare (1971) as the process by which a non-governmental organization grants recognition to an individual who has met predetermined qualifications specified by that organization. The National Organization for Competency Assurance (2006) defines certification as the voluntary process by which a non-governmental entity grants a time-limited recognition and use of a credential to an individual after verifying that he or she has met predetermined and standardized criteria. It is the vehicle that a profession or occupation uses to differentiate among its members, using standards, sometimes developed through a consensus-driven process, based on existing legal and psychometric requirements. While certification may be included as a component of the regulatory authority's requirement for issuance of a license, certification alone does not provide legal authority to practice within the profession. Similar to licensure, certification examinations are generally considered high stakes and require comparable standards related to evidence of validity.

Educational examinations (i.e., college admission examination, test preparation examinations, course progression examinations) are developed for fundamentally different purposes than licensure or certification examinations. Educational examinations are generally developed for the purpose of making judgments about the status, progress or

¹ Collaboration: Forging solutions through respect, diversity and the collective strength of all stakeholders.

Excellence: Striving to be and do the best.

Innovation: Embracing change as an opportunity to better all organizational endeavors and turning new ideas into action.

Integrity: Doing the right thing for the right reason through honest, informed, open and ethical dialogue.

Transparency: Demonstrating and expecting openness, clear communication, and accountability of processes and outcomes.

accomplishments of an individual related to a specified course of study or occupational discipline (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). An important difference between licensure examinations and educational examinations is the stated purpose. Licensure examinations primarily focus on protection of the public, while educational examinations focus primarily on individual student progression. Because educational examinations are developed to assist with a variety of judgments related to curricular progression, educational achievement and academic ability, the stakes of the examination may range from relatively low stakes to high stakes depending on the criticality of the judgment, thus allowing varying degrees of rigor with respect to technical quality and validity (AERA, APA, & NCME, 1999). Regardless of the stakes of the examination, an important difference between licensure and educational examinations is the scope of the content included within the examination. Because educational examinations measure educational achievement or progression, content decisions generally reflect a distinct set of curricular objectives, while licensure examination content decisions generally reflect a much broader set of knowledge, skills and abilities (KSAs) necessary for competent practice as defined by an entry-level practice analysis.

Introduction to the NCLEX® Computerized Adaptive Testing (CAT) Experience

The NCLEX examination is different than a traditional pencil-and-paper examination. Typically, pencil-and-paper examinations administer the same items to every candidate, thus ensuring that the difficulty of the examination is the same across the board. Because the difficulty of the examination is constant, the percentage correct is the indicator of the candidate's ability. One disadvantage of this approach is that it is inefficient. It requires the high-ability candidates to answer all the easy items on the examination, which provides very little information about their ability. Another disadvantage is that guessing can artificially inflate the scores of low-ability candidates because they can answer these items correctly 25 percent of the time for reasons that have nothing to do with their ability. Instead, the NCLEX examination uses CAT to administer the items. CAT is able to produce test results that are more stable using fewer items by targeting items to the candidate's ability. The computer's goal during the NCLEX examination is to determine the ability of the candidate in relation to the passing standard. Every time the candidate answers an item, the computer re-estimates the candidate's ability. With each additional item answered, the ability estimate becomes more precise.

Each item that the candidate receives is selected from a large pool of items using three criteria:

1. The item is limited to a content area that will produce the best match to the test plan percentages. It ensures that each candidate's examination has enough questions from each content area to match the required test plan percentages.
2. An item is selected that the candidate is expected to find challenging. Based on the candidate's answers up to that point and the difficulty of those items, the computer estimates the candidate's ability and selects an item that the candidate should have a 50 percent chance of answering correctly. This way, the next item should not be too easy or too hard and the computer can get maximum information about the candidate's ability from the item.
3. Excludes any item that a repeat candidate has seen in the last year.

For CAT to work, the difficulty of each item must be known in advance. The degree of difficulty is determined by administering the items as pretest items to a large sample of NCLEX candidates. Because the difficulty of these pretest items is not known in advance, these items are not included when estimating the candidate's ability or making pass/fail decisions. When enough responses are collected, the pretest items are statistically analyzed and calibrated. If the pretest items meet the NCLEX statistical standards, they can be administered in future examinations as scored items.

The decision as to whether a candidate passes or fails the NCLEX-RN and NCLEX-PN examination is governed by three different scenarios:

Scenario #1: 95 Percent Confidence Interval Rule

This scenario is the most common for NCLEX examination candidates. The computer will stop administering items when it is 95 percent certain that the candidate's ability is either clearly above or clearly below the passing standard.

Scenario #2: Maximum-Length Examination Rule

Some candidate's ability levels will be very close to the passing standard. When this is the case, the computer continues to administer questions until the maximum number of items is reached. At this point, the computer disregards the 95 percent confidence rule and considers only the final ability estimate.

- If the final ability estimate is above the passing standard, the candidate passes.
- If the final ability estimate is at or below the passing standard, the candidate fails.

Scenario #3: Run-Out-of-Time (R.O.O.T.) Rule

If a candidate runs out of time before reaching the maximum number of items and the computer has not determined with 95 percent certainty whether the candidate has passed or failed, an alternate criteria is used.

- If the candidate has not answered the minimum number of required items, the candidate automatically fails.
- If at least the minimum number of required items were answered, the computer looks at the last 60 ability estimates:
 - If the last 60 ability estimates were consistently above the passing standard, the candidate passes.
 - If the candidate's ability estimate drops below the passing standard even once over the last 60 items, the examinee fails. This does not mean that the candidate must answer the last 60 items correctly. Each ability estimate is based upon all previous items answered.

Candidates may be administered multiple choice items, as well as items written in alternate formats. These formats may include but are not limited to multiple response, fill-in-the-blank calculation, ordered response and/or hot spots. All item types may include multimedia, such as charts, tables, graphics and sound.

Exhibiting the Core Values in Ensuring Test Validity

Because the NCLEX examination is a major component for entry into nurse practice, it is considered to be a high-stakes examination and therefore necessitates high standards to ensure the examination is valid. It must demonstrate both validity and reliability in order to allow BONs to make defensible licensure decisions. The validity of a licensure examination depends on both its ability to measure competencies necessary for safe and effective practice and to distinguish between candidates who possess these competencies from those who do not. The reliability of a licensure examination is its ability to yield consistent results; to pass or fail candidates possessing the same level of competency consistently. It is ultimately the core values of NCSBN that require the establishment of the validity of the examination and guide its maintenance over time. As demonstrated in the remaining sections of this article, the elements of collaboration, integrity, excellence and transparency are the building blocks upon which the most important aspect of the NCLEX examination, validity, is built and maintained.

Historically, validity has been defined as the degree of accuracy of a measure. A valid measure assesses all relevant aspects of a trait and only that trait. Messick (1989) states that the major concern of validity is not to explain any single isolated event or response to an item because these almost certainly reflect a confounding of multiple determinants; rather, the intent is to account for consistency in behaviors or item responses, which frequently reflects distinguishable determinants. (p. 14) Since the 1950s, psychometric research and reporting has referred to three major categories of validity: content-related, criterion-related and construct related.

More recently, the AERA, APA, and NCME have developed a unified concept of validity using an argument-based approach as noted in the *Standards for Educational and Psychological Testing* (1999):

A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses... Ultimately, the validity of an intended interpretation...relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction, adequate score reliability, appropriate test administration and scoring, accurate score scaling, equating, and standard setting, and careful attention to fairness for all examinees. (p. 17)

As can be seen from this quotation, the *Standards for Educational and Psychological Testing* do not specify what type of evidence should be used to support that a licensure examination is measuring what is intended; rather, a validity argument is presented and evidence is accumulated to support intended interpretations and valid assumptions in major testing areas such as scoring, test administration, test content, standard setting and the interpretation/ consequences of testing.

The *Standards for Educational and Psychological Testing* specify that licensure examinations should meet certain criteria which demonstrates their validity, as well as adhering to the intent of state and federal laws. NCSBN, as a test-producing entity, must be able to show that the NCLEX examination assesses candidates on the essential KSAs required to practice nursing upon entry into practice. A license is seen by the public as evidence that the practitioner possesses the necessary KSAs to provide safe and effective professional services; thus the licensing entity should ideally be independent from the professional and/or educational arm of the profession in order to provide independent validity evidence to the public (Clauser & Margolis, 2006). BONs meet this requirement of independence. Furthermore, licensing entities, such as BONs, are expected to establish requirements that are sufficiently extensive and demanding to meet the public's expectations. They accomplish this by permitting candidates to take the licensure examination after completing an approved nursing education program.

The validation process involves the development of an argument in support of the interpretations and inferences drawn from examination scores and evidence to support those arguments. "The interpretative argument can be represented as a chain or network of inferences leading from examination scores to conclusions to be drawn and decisions based on those conclusions" (Kane, 1992). The interpretative argument provides an explicit statement regarding the inferences and assumptions inherent in the argument and provides a framework for evaluating the proposed interpretation. All inferences and assumptions must be sound if the argument is to be considered valid. Moreover, an evaluation of plausible alternative interpretations should be considered as part of the validity argument (Kane, 2006).

The general validity assumption regarding the NCLEX examination is that the examination measures whether or not a candidate has the requisite KSAs needed for entry into nurse practice. This general validity assumption is supported by evidence scoring, generalizability, extrapolation and interpretation.

The primary validity argument of the NCLEX examination is provided by the rigorous processes in place for CAT. To ensure that the score key is reasonable, great care is taken to establish that the items being administered to candidates are fair and reasonable. Before each item is administered to nurse candidates, it goes through many reviews to ensure the item has only one correct answer and meets prescribed statistical criteria. Additionally, all items being administered to candidates are reviewed just prior to deployment of an item pool to make certain that the item keys are correct. All items in an item pool undergo a preliminary item analysis shortly after the pool is released in order to confirm that the item is meeting statistical criteria and there is only one correct answer. Further statistical analyses, such as making sure that the data fit the Rasch model, are performed.

Additional scoring validity is provided by the use of standardized procedures and conditions using professional testing centers. For the NCLEX, prior to an item pool rotating into the field, a cross functional group of professional staff meet to check that the item pool is being deployed correctly and that the item selection algorithm is working correctly. Accurate application of the scoring rules of an NCLEX examination is ensured by scoring each candidate's examination twice prior to the release of results, as well as other quality control procedures. Evidence to support this assumption includes consistency in the administration of the examination through the use of professional test centers that ensure registration, administration and scoring of NCLEX examinations are standardized across the U.S.

A great deal of time and effort is spent guaranteeing that the examination is administered under standardized conditions. One of the compelling reasons for transitioning NCLEX to a CAT examination was the structured testing environment with close monitoring by professionals and up-to-date technology resulting in few variations in testing conditions. Any variations in testing conditions are documented for NCSBN and can be used to invalidate a score, and if necessary, to ensure the scoring rule is applied consistently for all candidates. These are just a sample of the data and quality control procedures that provide evidence to support this assumption and the interpretive argument that the rules used to score NCLEX candidates are appropriate and applied consistently and accurately.

The validity argument for generalizability is "reliability or generalizability studies as well as judgments about the representative of the sample of observations included in the test" (Kane, 2006, p. 25). For the NCLEX, decision consistency is used as an index of reliability. Decision consistency represents the proportion of pass/fail decisions that would remain the same if all candidates were to be re-tested with parallel examinations. Regarding the representativeness of the sample of observations (i.e., the items administered to a candidate) at least four times a year, subject matter experts (SMEs) review representative NCLEX examinations to provide evidence that the items are relevant and the examination possesses face validity.

Once a year, SMEs who are also familiar with nursing regulation, review representative examinations and provide expert opinion as to whether the examinations are “valid” for licensure decision. Additionally, the automated item selection algorithm, as well as the highly structured content and statistical specifications for the examination, provides evidence to support the generalizability of scores (Clauser & Margolis, p. 717). Moreover, if there is a fit between the data (candidate responses to items) and the Rasch measurement model, it is assumed that candidate’s performance on items on the NCLEX represents their ability on the NCLEX ability scale (Kane, 2006, p. 141). For the NCLEX, model-data fit is continuously monitored and provides evidence to support the generalization assumptions.

The validity argument for extrapolation suggests inferences from universal score to target sample (NCLEX candidates) are based on judgments about the overlap between the skills measured by the test and those needed to practice nursing upon entry into practice (Kane, 2006, p. 147). In general, evidence to support this assumption is provided by the rigorous test development activities and the content of the examination. The NCLEX test plan specifications are based on in-depth and comprehensive empirical studies and validated by stakeholders (Wendt & Kenny, 2007). All NCLEX items are coded to the job tasks identified in the practice analysis studies, which provide evidence that the items on the examination are based on the skills needed for entry into practice. Face validity reports confirm that the NCLEX examinations are relevant and appropriate for licensure decisions and there is appropriate representation on the examination of the skills needed to practice (i.e., neither over nor under representation of skills). Moreover, all items undergo an independent review by SMEs and during that process items are checked for relevance to practice.

NCSBN has processes in place to provide evidence that there are no skill irrelevant sources of variability that could threaten the interpretation of examination results. For example, the reading level of examinations are monitored and maintained at a minimum level so the examination measures nursing ability and not reading ability (Woo, Wendt & Liu, 2009). Fairness and sensitivity reviews are conducted to remove any construct irrelevant information from items, as well as to remove any irrelevant information in items that may be insensitive (Wendt, Kenny & Riley, 2009). Finally, differential item functioning (DIF) analyses are conducted on all NCLEX items with sufficient sample size in order to identify statistical indices of items functioning differently for various groups. Any items flagged for DIF are reviewed by SMEs and removed if there is construct irrelevant information (Wendt & Worcester, 2002). The NCLEX development processes are quite thorough and comprehensive, and help to provide evidence that the skills being assessed are those needed to practice, not irrelevant factors.

Decisions based on examination scores make assumptions about various desired outcomes and interpretations. Therefore, NCSBN conducts standard setting workshops where a criterion-referenced passing standard is recommended using a carefully selected panel of judges to provide evidence that the decisions regarding pass/fail performance are fair and not dependent on the group of candidates testing at a particular time. After a standard setting workshop, the NCSBN Board of Directors (BOD) reviews the recommended passing standard, multiple data sources, as well as impact data to make an informed decision regarding NCLEX passing standards. Standard setting takes place at least every three years or when a test plan is changed in order to provide evidence in support of a fair standard and that the decisions are appropriate (Wendt & Kenny, 2007). There are multiple quality control checks to ensure that the passing standard is applied consistently and the evidence is documented in internal documents and cross-functional meetings.

Regarding the interpretation of licensure examination results, there is some variation in what is reported to stakeholders, such as a score, scaled score or pass/fail. In general most interpretations of the results will certify whether or not the candidate is suitable to be licensed and to practice. The purpose of the NCLEX using CAT is well documented and evidence to support the interpretation of it is accepted. No argument is made that the examination helps to rank order candidates. Thus, examination results are reported as pass or fail rather than a score where the results could have the potential to be misused for job selections. Candidates are allowed to retake the NCLEX after 45 test free days in many jurisdictions, thus access to practice is not restricted other than by the candidate’s ability. More importantly, the purpose of the licensure examination is to protect the public and that goal is accomplished with the score interpretation that those who pass the NCLEX have the KSAs needed for entry into practice.

Validity arguments relative to the NCLEX examination are not complete until other alternative interpretations are presented and rejected. Any break in examination security is considered a possible threat to all other validity arguments. One type of security threat is a proxy test taker of the NCLEX examination. Should this occur, one of the assumptions that a score represents a candidate’s true ability would be false. For the NCLEX examination this proxy

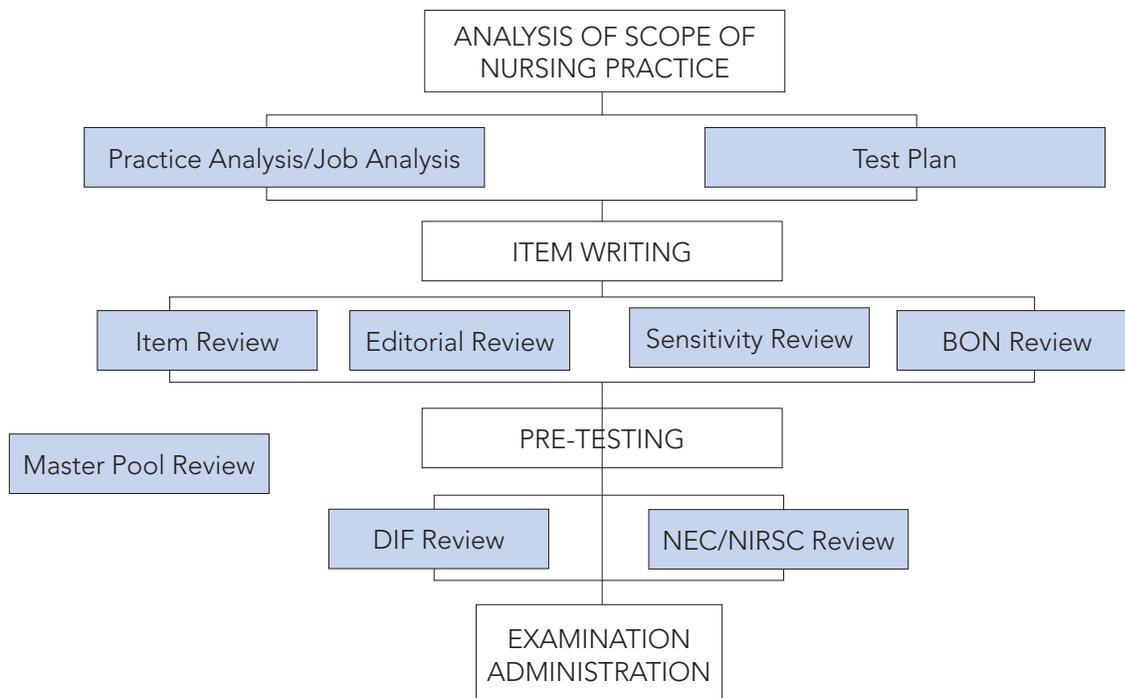
test taker scenario is not likely to happen because of the many security procedures in place. For the NCLEX, these security procedures involve strict identification procedures, including the candidate's authorization to test (ATT), fingerprinting, palm vein scanning, as well as monitoring at professional test centers. This particular threat to a validity interpretation seems minimal.

A more plausible security threat to our validity argument is if candidates have advanced knowledge of NCLEX items. To mitigate this security risk, NCSBN uses two security firms to search for possible security risks and item content on the Internet. High risk sites are identified and investigated. Any items that are deemed compromised by SMEs are removed from administration. Research has been conducted regarding the memorability of NCLEX items (Wendt & Harmes, 2009). Results indicate that it would take a very organized effort to steal sufficient numbers of NCLEX items for candidates to memorize those items and for those items to be administered to candidates. Nevertheless, NCSBN remains vigilant in its effort to closely monitor test and Internet sites for statistical indications of possible security breaks. When necessary, investigations are conducted and steps are taken to ensure the security of the NCLEX and the interpretative arguments.

Another threat to the NCLEX validity argument stems from threats that performance on examination items may be influenced by factors other than what is intended by the examination items. For example, it could be that the use of the computer interface to take a CAT test introduces factors that are not associated with nursing ability. Clearly, the beta testing of CAT for NCLEX has provided evidence that this is not the case. In this same line of thinking, use of the computer could increase test anxiety for some candidates. However, the use of computers for testing is extremely common and more importantly, nurses in practice need to use the computer to access information. Furthermore, a tutorial that allows candidates to practice using the computer interface is available on the NCSBN website and each candidate must take the same computer tutorial at the beginning of their testing session. Thus, computer skill and a potential of anxiety related to its use in testing is not likely to be a plausible threat to the validity argument.

Exhibiting the Core Values in the Examination Development Process

NCSBN depends on currently practicing nurses and nurse regulators to assist in the NCLEX item development process. This process is a key component in maintaining high quality NCLEX examination items. The NCLEX examination must be reliable, valid, psychometrically sound and legally defensible. The NCLEX development process is a four-step process, including analysis of the scope of practice, item writing, pretesting and administration, as represented in the chart below.



The NCLEX examination development process is an exemplar of collaboration, excellence, integrity innovation, and transparency. The examination development process includes several rigorous steps requiring the participation of thousands of individuals, including nurses, nurse regulators, examination development experts and psychometricians. Adhering to the quality guidelines published in the *Standards for Educational and Psychological Testing* (1999), the first step in every NCLEX examination development cycle begins with analysis of the current scope of practice. Ensuring representation from the four nursing regions of the U.S.², five to 10 SMEs representing a variety of nursing practice settings and geographic diversity are chosen to provide input as members of the practice analysis panel. These SMEs, using professional experience, observation of entry-level nurses in various practice settings, knowledge of nursing practice acts and entry-level nurse orientation manuals, develop survey instruments designed to determine how frequently entry-level nurses (defined as six months or less post-licensure) perform essential nursing activities and how important those activities are to safe and effective entry-level practice. Although not required in the *Standards for Educational and Psychological Testing* (AERA, 1999), NCSBN also conducts a KSA analysis prior to developing each examination. The KSA panel is selected using the same procedures for the practice analysis panel selection. The KSA panel is charged with developing a survey instrument designed to identify the necessary KSAs needed by entry-level nurses to safely and effectively perform the activities identified in the practice analysis survey instrument. The separate survey instruments developed by the panels are then sent to a random selection of more than 12,000 entry-level nurses. This entire process involves the collaboration of survey design experts, psychometricians, nurse supervisors, nurse clinicians and several thousand entry-level nurses across the U.S.

Following the completion of the practice analysis and KSA studies, the NCLEX® Examination Committee (NEC), appointed by the NCSBN BOD and comprised of 10 nurses with extensive experience in nursing and nursing regulation, is charged with compiling the evidence obtained from the studies and recommending the most appropriate content distribution and entry-level nursing activities to include on the NCLEX test plan. The NEC's recommendations are then sent to all of the BONs, along with supporting evidence-based documents, for feedback and comment. After considering the feedback provided by the BONs, the NEC finalizes its recommendation for the NCLEX test plan and presents it to the NCSBN Delegate Assembly for approval. Using evidence-based outcomes and nurse regulatory input during this phase of analysis of scope of practice promotes collaboration of testing specialists with nurse regulatory expertise to ensure transparency and integrity of the examination development process are rigorously maintained.

The next phase of the examination development process begins with item (examination question) construction. NCSBN maintains a large database of volunteer nurse item writers, appointed by their BON, and representing a variety of practice and educational settings. After completing an item needs analysis, six to eight nurses are selected from the database and invited to attend a four-day item writing panel. During fiscal year 2010 (FY10), more than 141 volunteers participated as members of NCLEX item writing panels. In order to help ensure the quality and excellence of the items constructed, panel members were provided with extensive training in high stakes item writing principles and NCLEX item writing style guide. Each panelist is required to provide evidentiary validation of item content and the correct answer using commonly available nursing textbooks, journals and practice guidelines. Every item developed by the panel is subsequently forwarded to the nursing content and professional editorial staff of NCSBN test services provider for review. These individuals evaluate and edit each item to ensure the content addresses the activities identified in the NCLEX test plan, adheres to the NCLEX style guide, and that the validations of content and correct answer are accurate.

While this process meets the minimum standards for quality item construction, in keeping with the core value of excellence, the items are then reviewed by an additional independent group of expert nurses selected from the NCSBN database of nurse volunteers. These volunteers are item reviewers, nurses who work with entry-level nurses as supervisors, mentors and preceptors. Like the nurse item writers, these nurses are approved by their BONs and attend a four-day meeting as item review panelists. They are charged with reviewing items to ensure items reflect

2 Area 1: Alaska, American Samoa (AS), Arizona, California, Colorado, Guam (GU), Hawaii, Idaho, Montana, Nevada, New Mexico, Northern Mariana Islands (MP), Oregon, Utah, Washington, Wyoming

Area 2: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, West Virginia, Wisconsin

Area 3: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia

Area 4: Connecticut, Delaware, District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, U.S. Virgin Islands, Vermont

current practice and that items with regional variations/biases are corrected or removed from use in potential NCLEX examinations. All items surviving the item review process are returned to the NCSBN test service provider for a second editorial review.

Finally, prior to the pretesting phase of item construction, all items surviving the second editorial review are presented to a sensitivity panel. The sensitivity panel is composed of individuals from various ethnic backgrounds, including at least one nurse and one linguist. The sensitivity panel is responsible for reviewing all items to identify words and phrases that might be viewed as insensitive, stereotypical, or inflammatory to a specific group of people or population. The pretesting phase of the examination development process is designed to gather statistical information on item performance after the item writing phase is complete. The items prepared for pretesting are included with the operational NCLEX items and administered to nursing candidates to help ensure statistical information gathered on the items are a true representation of responses provided by the population of entry-level nurses. Only item information gained from nurse candidates educated in the U.S. and attempting the examination for the first time are included in the pretest statistical information. In addition, a minimum of 400 candidates meeting the inclusion criteria must respond to an item before its statistical information can be evaluated. The required sample size and inclusion criteria have been established to help ensure the reliability, validity and generalizability of the statistical information while promoting the excellence and integrity of the examination development process.

Pretest items meeting the NCSBN stringent statistical criteria related to difficulty, discrimination and distractor response patterns are once again subjected to an independent review process. The NCLEX® Item Review Subcommittee (NIRSC), whose members are nominated by BONs and appointed by the NCSBN BOD, is composed of nurses who have comprehensive knowledge of their respective nurse practice acts and practice of entry-level nurses. These nurses are charged with reviewing the items for construction flaws, correct validation citations, compliance with entry-level practice and violations of nurse practice acts. The NIRSC may return flawed items to the beginning of the item writing process for rework, forward items in which the committee could not reach consensus to the NEC for further review or approve items for use in future NCLEX operational item pools.

Additional Examination Development Processes in Support of Core Values

DIF

Once NCLEX items have been approved for use in NCLEX operational item pools, the core values of collaboration, excellence, integrity, innovation, and transparency continue to guide the validity and technical quality of the examination items through the activities of the DIF Committee, Member Board Reviews, and Master Pool Review Panel.

Once an NCLEX operational pool is deployed and sufficient candidate responses are obtained, all items are statistically assessed for DIF. DIF refers to the potential of examination items to behave differently with respect to item parameters (i.e., difficulty, discrimination, guessing) when administered to different subgroups of a specified population who have been matched on the ability purported to be measured by the examination. DIF, an undesirable characteristic of an examination, implies the item is measuring the construct for which the examination was developed (e.g., nurse entry-level competence), as well as another characteristic that is solely dependent on membership within a subgroup (e.g., gender, race, ethnicity) of the target population.

The principles of test fairness, as well as the core values of excellence and integrity, require that NCLEX operational items undergo scrutiny to detect and remove items that behave in significantly different ways for different groups solely based on these types of demographic characteristics. The terms "reference group" and "focal group" are used in DIF for group comparisons and refer generally to the majority (reference group) and the minority (focal group) for the examination population. Of particular importance in understanding the concepts of DIF is the requirement that reference and focal groups be matched on ability before evaluating for differences in item behavior. Any significant differences in item behavior after the subgroup population has been matched on ability level of the reference group should be flagged for DIF and evaluated for item bias. For example, if a male and a female NCLEX candidate of the same ability level perform differently on an item, the item may be measuring something other than the ability of the candidates. In this example, it is reasonable to consider the possibility that the item is actually measuring some aspect of the candidate's ability related specifically to gender.

Supporting the core values of integrity, excellence and transparency, NCSBN requires that each experimental and operational pool be examined for potential DIF. Every item occurring in an experimental or operational pool with at least 50 focal group candidate responses and at least 400 reference group candidate responses is included in the DIF analysis to ensure items do not contain gender or ethnicity bias. Items that have a statistically significant difference using a t-test with significance $\leq .0001$ and a difference in magnitude between focal and reference groups ≥ 0.50 logits are identified as possessing potential DIF. These items are then forwarded to the NCLEX® DIF Review Panel for review. The NCLEX® DIF Review Panel consists of a minimum of five members, including at least one male and at least three of the six ethnic focal groups included in the DIF analysis. No less than one individual must have prior experience on a DIF review panel, and at least one individual must have a linguistic background. In addition, one member of the panel must be a registered nurse (RN) with a current license. The NCLEX® DIF Review Panel is charged with reviewing all items to determine if the item possesses true bias toward any ethnicity or gender. Items determined to possess true bias are referred to the NEC for final disposition.

Member Board Reviews

Member Board Reviews represent a collaborative effort between NCSBN and BONs to help ensure that only items meeting the nurse practice act of all BONs remain in the NCLEX operational item pool. Twice a year, each BON is permitted to review items in the current operational pool for the purpose of identifying any item that would not be supported by their specific nurse practice act. Items identified as problematic by the BON are forwarded to the NEC for final disposition.

Master Pool Review

Major evidence for validity and technical quality of the NCLEX examination are established by the examination development process because nurse practice evolves over time based on nursing research, client complexity and workforce issues; the validity and the technical quality of the items measuring nurse competence must be maintained over time. Therefore, in support of the integrity and excellence of the examination, every item in the NCLEX item bank is required to be reviewed every four years. This review is completed by a panel of nurses from the four nursing regions with demonstrated expertise as a nurse and prior experience as an item reviewer. During this review all items are reviewed for currency to practice, item construction flaws and currency of evidentiary validation.

Exhibiting the Core Values through the Standard Setting Process

The standard setting process is another example of the core values of NCSBN in action, supporting collaboration, evidence-based decisions and excellence in upholding integrity of the examination scoring process. The standard-setting process utilizes information from a variety of stakeholders in nursing and testing, including nurses, educators, employers, regulators and psychometricians. The passing standards for the NCLEX examinations are re-evaluated by the NCSBN BOD every three years. Standard setting is a process that requires the BOD to consider a variety of evidence in order to establish an examination passing standard which best ensures that NCLEX candidates are competent and able to practice safely and effectively.

The BOD considers the results of an annual survey sent to nursing educators and nursing employers regarding the competence of the current group of entry-level nurses. Survey respondents are asked to provide feedback on the level of competence of entry-level nurses and whether the current NCLEX passing standard is set at an appropriate level. Surveys for licensed practical/vocational nurse (LPN/VN) and RN employers and educators are administered separately. A panel of judges is also convened to conduct a standard setting workshop. The panel of judges utilizes the Angoff method (1984) supplemented by the Buek (1984) in order to provide a passing standard recommendation to the BOD. Additionally, the BOD reviews the outcomes of previous passing standard setting decisions, past and current data describing NCLEX candidate performance, and national data reflecting graduating seniors' preparedness for entry into the nursing profession. After full consideration of all evidence, including their personal knowledge about client acuity, nursing education and nursing practice, the BOD makes an informed decision relating to a reasonable and legally defensible passing standard that most closely represents the minimal ability level necessary for safe and effective nursing practice.

Introduction to the NCLEX® Administration Experience for the Nurse Candidate

NCLEX candidates begin the process of taking the NCLEX by registering for the examination with NCSBN's test service provider. At this time the candidate also needs to apply for licensure with the BON where they wish to practice in order to be made eligible to schedule an NCLEX appointment. Once the BON applies eligibility to the NCLEX registration with the test service provider the candidate receives an ATT letter and can schedule an examination appointment.

The examination appointment begins with a check-in processes, which includes reading and signing the NCLEX® Examination Candidate Rules agreement; providing biometrics (fingerprint, palm vein scan and having a photograph taken); and providing acceptable identification and ATT letter. The candidate is then given an erasable note board and pen before being seated in the examination room. The candidate begins the examination session by completing a short tutorial before answering actual NCLEX items. When the examination is finished, the candidate is asked to answer a short survey about their examination experience. During the examination, candidates are video and audio recorded at all times.

NCLEX results are double-checked and then transmitted back to the BON; the BON mails the candidate their official results within four weeks of the examination appointment. Some BONs allow candidates to check their unofficial results through the quick results service. Candidates who have failed the examination receive a candidate performance report summarizing their performance on each of the NCLEX test plan content areas in order to assist them in preparing for their next examination. To retake the examination, candidates will have to reregister with Pearson VUE and determine if anything additional is required by the BON in order to receive eligibility again. Candidates will receive a new ATT letter and can schedule no sooner than 45 days from their previous appointment; this retake time frame is determined by the BON.

Exhibiting the Core Values in the Examination Administration Process

The NCSBN Examinations department works closely with its test service provider and BONs to ensure that the administration of the NCLEX examination is an effort of collaboration. The NCLEX® Administration Website allows NCSBN to collaborate with BONs and the test service provider to manage candidate information, and to address and resolve issues related to examination accommodations, test security, and examination registration issues. Additionally, NCSBN works directly with BONs, the test service provider and leaders in the testing industry to evaluate and improve testing processes, including, but certainly not limited to, identification requirements, security, incident reports and American Disabilities Act (ADA) testing accommodations. The NCLEX® Administration Website also allows complete transparency of key performance indicators of the examination administration processes by providing BONs direct and immediate access to data and reports about examination performance and examination activities within their jurisdiction. Collaboration and transparency are again demonstrated by the regular reviews and revisions of the NCLEX examinations website and the NCLEX® Administration Website in order to provide up-to-the-minute, clear information to the public and stakeholders, as well as providing more advanced reporting options for BONs.

The *NCLEX® Member Board Manual* documents policies and procedure that act as the foundation for standardization of the administration of the NCLEX examination. In an effort to remain current in all aspects of examination development and administration, NCSBN maintains contact with other small and large scale examination programs in order to contribute knowledge to the industry, and gain insight on innovations and industry standards that enhance NCLEX examination administration (e.g., security processes and procedures). As evidence of this collaboration with testing industry professionals, NCLEX, in partnership with the test service provider, began using palm vein scan technology to improve its biometric security measures at each test center. Through collaboration with its test service provider, NCSBN was also able to implement an update test driver into the CAT technology that allowed for the implementation of more complex alternate item types.

As discussed earlier, any breach of security is a threat to the validity and integrity of the examination. NCSBN administration works with multiple Web patrol services to ensure that NCLEX information remains secure and is not shared amongst public parties. Stringent test center admittance policies and procedures, as well as a staff of trained and certified test administrators, add an additional layer to protecting the integrity of the NCLEX examinations.

NCSBN Examinations strives to provide transparency between the NCLEX program, BONs, nurse candidates and the public through the use of an open communication plan. Through social media outlets, NCLEX maintains direct

and timely contact with NCLEX candidates and educators, answering questions and providing the most important information they need to be successful, including policies and procedures they need in order keep them aware of the rules surrounding the important, high-stakes nature of NCLEX information. Additionally, NCSBN Examinations communicates with BONs on a daily basis to discuss and resolve administration issues (e.g., ADA accommodations). BONs are also provided with a complete communication package, including the annual NCLEX Conference, topic-specific webinars and publications like the *NCLEX® Examination Candidate Bulletin* and the *NCLEX® Member Board Manual*.

Conclusion

NCSBN's adheres to the core values of collaboration, excellence, innovation, integrity and transparency in all aspects of support of its mission. As demonstrated by this article, the elements of examination development and administration of the NCLEX examinations are exemplars of the core values of NCSBN in action. The fundamental validity arguments related to the NCLEX examinations are directly supported by the transparency of the item development process, the collaboration required in the development of activity statement through the practice analysis process, innovation through development of alternate item styles, and the integrity and excellence principles adhered to during the standard setting process. The rigorous administration and security procedures of the NCLEX examination are additional examples of the use of NCSBN core values to support regulatory excellence.

References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington DC: American Psychological Association.
- 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.
- Clauser, B. & Margolis, M. (2006). Testing for licensure and certification in the profession. In R. Brennan (Ed.), *Educational Measurement* (4th Edition, pp. 702-735). Westport, CT: Praeger Publishers.
- Impara, J. C. (Ed.). (1995). *Licensure testing: Purposes, procedures, and practices*. Lincoln, NE: Buros Institute of Mental Measurements.
- Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112, 527-535.
- Kane, M. (2006). Content-related validity evidence in test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 131-154). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*. (3rd Edition, pp.13-103). New York: Macmillan.
- National Organization for Competency Assurance (NOCA). (2006). *NOCA's basic guide to credentialing terminology*. Washington DC: Author.
- O'Neill, T., Marks, C., & Reynolds, M. (2005). Re-evaluating the NCLEX-RN passing standard. *Journal of Nursing Measurement*, 13(2), 147-165.
- Schoon, C. G. & Smith, I. L. (Eds.). (2000). *The licensure and certification mission*. New York, NY: Professional Examination Service.
- U.S. Department of Health, Education, and Welfare. (1971). *Report on licensure and related health personnel credentialing*. Washington, DC: Author.

- Wendt, A. & Worcester, P. (2000). The NCLEX-DIF process. *Journal of Nursing Education*, 39(4), 185-187.
- Wendt, A. & Kenny, L. (2007). Setting the passing standard for the national council licensure examination for registered nurses. *Nurse Educator*, 32(3), 104-108.
- Wendt, A., Kenny, L., & Riley, M. (2009). Fairness and sensitivity review. *Nurse Educator*, 34(5), 228-231.
- Wendt, A. & Harnes, C. (2009). Memorability of innovative items. *CLEAR Exam Review*, 20(1), 16-20.
- Woo, A., Wendt, A., & Liu, W. (2009). Readability of licensure examinations. *CLEAR Exam Review*, 20(1), 21-23.

©2011 The National Council of State Boards of Nursing (NCSBN) is a not-for-profit organization whose members include the boards of nursing in the 50 states, the District of Columbia and four U.S. territories — American Samoa, Guam, Northern Mariana Islands and the Virgin Islands. There are also nine associate members.

Mission: NCSBN provides education, service and research through collaborative leadership to promote evidence-based regulatory excellence for patient safety and public protection.

 **NCSBN**
National Council of State Boards of Nursing
111 E. Wacker Drive, Ste. 2900
Chicago, IL 60601
Main 312.525.3600
Fax 312.279.1032
www.ncsbn.org