National League for Nursing

THE FAIR TESTING IMPERATIVE IN NURSING EDUCATION

A LIVING DOCUMENT FROM THE NATIONAL LEAGUE FOR NURSING

**NLN BOARD OF GOVERNORS
FEBRUARY 2012; UPDATED NOVEMBER 2020**

THE FAIR TESTING IMPERATIVE IN NURSING EDUCATION

A LIVING DOCUMENT FROM THE NATIONAL LEAGUE FOR NURSING

INTRODUCTION

The National League for Nursing recognizes the pressure faced by nursing programs to maintain strong NCLEX-RN® or NCLEX-PN® pass rates. While some stakeholders are concerned primarily with issues of public safety, reputation, recruitment of qualified students, competition, and accreditation, faculty and students are perhaps the parties most invested in student success – in the program and, eventually, on the NCLEX. No substantive debate has occurred about the need to protect the public through high-stakes assessment of competence using measures with robust validity and reliability evidence, such as the NCLEX-RN and NCLEX-PN exams. The American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (AERA et al., 2014), in their fifth edition of the *Standards for Educational and Psychological Testing*, define a high-stakes test as, "…used to provide results that have important, direct consequences for individuals, programs, or institutions involved in the testing" (p. 230).

While credentialing and licensure exams are widely acknowledged as examples of high-stakes exams, the term *high-stakes* is in reference to nursing education achievement exams when scores from these commercially available exams are used to make important educational decisions, such as whether students may continue to progress in their programs of study or whether they are permitted to graduate, pursue licensure, and enter the workforce. *It is the high-stakes use of commercially available, standardized nursing achievement exams, and specifically those exams designed as "exit exams" or "NCLEX predictor" exams, to prevent program completion, graduation, or in some other way deny eligibility to take the licensing exam – that is most concerning to the NLN and is, therefore, the subject of this Vision Statement.*

To address high-stakes testing in nursing education, the NLN undertook a process in 2010 to develop national fair testing guidelines to assist nurse faculty and administrators in creating and implementing ethical and evidence-based academic progression and graduation policies. The current document is an update to the 2012 NLN *Vision Statement, The Fair Testing Imperative in Nursing Education*. The fields of measurement and assessment are rapidly developing areas and, in response, updated *Standards* were published in 2014 by AERA et al., the first update since 1999 for guidelines that have existed for over 50 years. The most substantial change in the 2014 *Standards* is its much stronger focus on fairness in testing. Fairness has its own chapter and is weaved throughout the document, including fairness in test use and test score interpretation. To further illustrate, the 2014 *Standards* consider principles of test fairness to be equal in importance to the quality of the empirical evidence supporting a test's validity and reliability arguments (Plake & Wise, 2014).

BACKGROUND AND SIGNIFICANCE

In 2010, the NLN Presidential Task Force on High-Stakes Testing, leaders in nursing education, nursing practice, health care, and higher education, was called together to develop policy guidelines for the use of end-of-program testing, explore current thinking, and make recommendations for policies and practices in nursing education programs. These nursing-specific Fair Testing Guidelines were not meant to replace existing national standards but to contextualize key considerations from those standards for the profession of nursing.

Across the United States, a number of schools of nursing had implemented progression policies based on scores from commercially prepared, standardized nursing achievement exams (Spurlock, 2006). But, because no universally accepted standards existed for how predictive tests and related policies should be implemented, individual schools struggled to implement policies and standards on their own. To establish a baseline understanding of how commercially available high-stakes tests were being used as part of progression policies, the NLN collected descriptive data alongside the 2011 NLN Annual Survey. The results are summarized below:

For RN programs, initial findings indicated:

- Approximately one in three schools required pre-licensure RN students to obtain a minimum score on a standardized test in order to progress.
- Twenty percent of schools required a minimum test score to graduate.
- Twelve percent of schools did not forward students' names to state boards for licensure exam registration unless they reached minimum standardized test scores.
- Twelve percent of schools required that students meet minimum score levels at more than one point, or juncture, in the program.

Results also revealed that high-stakes testing was used only slightly less frequently in PN programs:

- Nearly one in four reported some use at least once in their curricula.
- Eighteen percent required a minimum test score to graduate.
- Thirteen percent did not forward students' names to state boards for licensure exam registration unless they reached minimum standardized test scores.

More recently, Randolph (2017) reported that administrators of 34 of Arizona's 38 prelicensure nursing education programs responded to a survey describing standardized testing practices. Ninety-seven percent of the programs were using commercially available standardized exams in primarily low- (but sometimes high-) stakes ways throughout the curriculum. Randolph compared schools that had established cut scores for the commercial exams with those that had not established cut scores (indicating a more formative than summative approach) and found no statistically significant differences in NCLEX-RN pass rates between the groups.

Certainly, standardized achievement exams are useful in numerous ways. They provide students with information about their knowledge compared to other students, using national and even program-level normative data. Standardized exams can also help faculty identify curricular strengths and weaknesses. But, requiring students to achieve a specific cut score on a commercially available standardized "exit

exam" in order to graduate and eventually sit for the NCLEX, in order to ensure that program pass rates remain at state board-prescribed levels, is especially problematic for those students who have successfully passed all other components of the nursing program but struggle to achieve the established cut score on the exit exam. Students who cannot achieve the required cut score may be required to retake the exit examination repeatedly until they achieve the score; in extreme cases, without having achieved the exit exam cut score established by the program, students may be denied their degrees or authorization to take the NCLEX. While in times past, separate and distinct progression policies based on exam cut scores were more common, an approach being increasingly used is for students not achieving the required cut score on the exit exam to fail the nursing course in which the exam is administered, endangering their standing in the nursing program. Such an approach can adversely affect students and their families economically, that is, while licensing is postponed, full salary potential is in jeopardy.

Students or groups of students who have suffered negative consequences for performing poorly on standardized tests have filed suit against their nursing programs using a variety of legal bases. Grounds for litigation may include breach of contract, lack of due process, and even educational malpractice, if standardized tests were inappropriately placed in the curriculum and utilized. High-stakes testing and progression/graduation policies can also distort the intended purpose of NCLEX pass rate requirements. First-time pass rates are viewed by stakeholders as a measure of program quality. Nursing programs that achieve high first-time pass rates by allowing only the highest performing students to sit for the licensing exam illustrate the well-known effect of selection bias.

The use of fair-testing principles may provide support to faculty as they grapple with these complex issues.

The Literature

A recurrent theme in the research literature on high-stakes testing in nursing is that most commercially available standardized predictive tests provide individual student scores that can accurately estimate a student's probability of *passing* the NCLEX-RN, and can easily classify students receiving high exit exam scores as likely to pass the licensure exam. Numerous studies have found that while predictive exit exams often work well in identifying high-performing students who are likely to pass the NCLEX, they are much less precise in estimating the probability of NCLEX failure (Brodersen & Mills, 2014; Hunsicker & Chitwood, 2018; Spurlock & Hunt, 2008; Spurlock, 2013). Drawing a distinction in describing the overall *accuracy* of test classifications, in which the accuracy of all exam predictions is figured, is especially important when policies that prevent progression or graduation are in place. Bowers and Zhou (2019) provide an exhaustive review of the relatively simple methods used to examine the classification accuracy of educational tests, and Smolkowski and Cummings (2015) provide an accessible introduction with example applications of relevance to nursing faculty.

Best practice guidelines indicate that the score a student must achieve to move forward in a program should be chosen 1) after a complete review of the test's psychometric parameters available from the test vendor; 2) after consideration of demographic, cultural, and linguistic issues among the student population; and 3) perhaps most importantly, after an assessment of the ethical and empirical basis for setting the cut score. One final consideration related to fairness in nursing as described by AERA et al.

(2014) is the consideration of unintended (and intended) consequences. While a program may wish to improve its NCLEX-RN pass rate quickly to avoid running afoul of the accreditor or state board of nursing, if the methods used to produce improvements in program pass rates result in, for example, a disproportionate negative effect on minority student progression and completion, then this unintended consequence must be tolerable to the program and satisfactory to the program's overseeing bodies.

The rigor and complexity of setting cut scores is more fully outlined by Zieky and Perie (2006); it is a process followed closely in setting the NCLEX-RN passing score (Wendt & Kenny, 2007). Because nursing programs are generally unequipped to undertake such a process, they may make cut score determinations in arbitrary ways, including relying only on the advice of the commercial test provider. Faculty must be especially guarded in their decision-making when little or no evidence exists to guide the setting of cut scores. The NLN's Fair Testing Guidelines and, to a much deeper extent, the *Standards for Educational and Psychological Testing* (AERA et al., 2014), provide sound advice for faculty navigating such weighty decisions. Studies have documented how standardized tests themselves show differential functioning for white vs. non-white examinees (Helms, 2005, 2006; Solórzano, 2008; Stout, 2002), especially minority students who often face language difficulties, social bias, stereotype threat, poor early academic preparation, and other difficulties brought about by structural inequities unconsciously built into routine processes and consciously enacted by unsupportive peers and faculty.

Despite the important role that NCLEX pass rates have played as the primary quality indicator for prelicensure nursing education for decades, recent research has revealed that more often than not, program-level factors – not student-level factors – best explain a program's NCLEX outcomes. Odom-Maryon et al. (2018) examined data from 832 prelicensure programs in the United States and found that, in order of importance, not requiring a standardized exam for admission, being housed in a public institution, and the percentage of full-time faculty were associated with program NCLEX pass rates. The use of standardized exit exams was not associated with NCLEX pass rates (OR = 1.22; 95% CI [0.84, 1.76]).

Similarly, Spector et al. (2020), in groundbreaking work, conducted an extensive literature review, a quantitative investigation of predictors of nursing program quality extracted from thousands of annual reports nursing programs submitted to their boards of nursing each year. They also conducted a Delphi study using experts in nursing education, regulation, and practice to identify other quality indicators of nursing education programs. Spector et al. noted that, despite how many faculty might characterize the relationship between NCLEX pass rates and program quality, declining NCLEX pass rates were clearly a downstream consequence of other problems – rather than being the primary problem itself. The quality indicators Spector et al. identified include those reflecting on the quality of school/program leadership and faculty support, the presence of consistent and competent faculty (with expertise in pedagogy), quality hands-on clinical experiences with trusted clinical learning partners, and an updated, evidence-based curriculum with a strong focus on patient safety, quality, and clinical reasoning. Because these factors were found to be early indicators of future NCLEX pass rate challenges (or strengths), Spector et al. have pointed researchers in nursing education in a perhaps somewhat more fruitful direction than the last several decades of focusing on student-level factors has proven to be.

THE NLN'S RESPONSE: FAIR TESTING GUIDELINES

The NLN has developed fair testing guidelines, based on the League's core values of caring, integrity, diversity, and excellence, and on widely accepted testing principles including, most notably, the 2014 *Standards for Educational and Psychological Testing* (AERA et al., 2014). Fairness, in this context, means that all test-takers are given comparable opportunities to demonstrate what they know and are able to do so in the learning area being tested. It also means that test scores, especially scores from high-stakes exams, must only be interpreted in ways supported by sufficient validity and reliability evidence. And finally, that the unanticipated consequences of high-stakes testing, and especially impacts on minority and other vulnerable students, should be considered during the construction of program policies guiding use of these exams and in the results obtained from the exams once in use.

The NLN Fair Testing Guidelines for Nursing Education value students' perspectives and backgrounds, and acknowledge the role of faculty in their implementation.

The Guidelines in Brief:

1) Faculty have an ethical obligation to ensure that both tests and the decisions based on tests are supported by solid validity and reliability evidence, are consistent across courses, and fair to all test takers regardless of age, gender, disability, race, ethnicity, national origin, religion, sexual orientation, linguistic background, testing style and ability, or other personal characteristics.
2) Faculty have the responsibility to assess students' abilities and assure that they are competent to practice nursing, while recognizing that current approaches to learning assessment are limited and imperfect.
3) Multiple sources of evidence are needed to evaluate basic nursing competence. Multiple approaches for assessment of knowledge and clinical abilities are particularly critical when high-stakes decisions (such as progression or graduation) may be made based on an analysis of these factors.
4) Tests and other evaluative measures should be used not only to evaluate student achievement, but, as importantly, to support student learning, improve teaching, and guide program improvements.
5) Comprehensive testing, administration, and evaluation information must be readily available to faculty before they administer, grade, and distribute results from, or write policies related to, the use of standardized tests. Faculty have the responsibility to review and understand the technical specifications of the standardized exams being used and to incorporate as much of this information as possible into communications with students about standardized testing and its consequences.

RECOMMENDATIONS

Based on the best available evidence on testing practices from nursing and other disciplines, the recommendations have been developed with a keen awareness of the pressure faculty face to meet local and national standards for program quality – and the tremendous obligation they feel to assess student ability and competence. Faculty are aware that evaluative measures should be used not only to evaluate student achievement, but, as importantly, to support student learning, and evaluate and improve teaching and program effectiveness.

The NLN recognizes the immense political, regulatory, and economic pressures placed on faculty and

programs to graduate classes with high first-time pass rates; *nurse educators face this pressure more than any other health care discipline*. The NLN has supported research addressing factors that are directly linked to a graduate's success in passing the NCLEX the first time. Though such factors are beyond the scope of this vision statement, the following recommendations are offered to strengthen academic policies that are formulated and implemented to address high-stakes testing.

Recommendations for Faculty

- In making decisions about using standardized tests, require written comprehensive information about the test, including evidence that the tests under consideration have been developed to minimize cultural and linguistic bias. Information about norms and norming procedures, reliability, and validity should also be reviewed by faculty before making a decision to adopt and use any standardized test.
- In developing a policy based on test results, include the core principle that multiple sources of evidence are fundamental to evaluate basic nursing competence. This is especially true when high-stakes decisions will be made based on the assemblage of numerous facts, including test scores.
- Teach students about the purpose of the tests, student-level factors that can affect the results, and the testing methods used within a program. Ensure that students receive information about precautions to take if potential employers request their standardized test results.
- Design and conduct research on fair testing practices and the impact on student diversity, learning, progression, and program completion.

Recommendations for Deans, Directors, Chairs

- Provide appropriate resources for faculty and staff to develop knowledge and skills in using multiple approaches to assessing student learning and nursing competence.
- With faculty, review standardized testing practices and progression policies based on test results to ensure alignment with accepted fair testing practices.
- Provide leadership in creating a climate based on your core values to support quality improvement using a learning assessment process. This includes standardized test results that can be aggregated, trended, and analyzed.

Recommendations for the National League for Nursing

- Regularly update the NLN Fair Testing Guidelines for Nursing Education through the NLN's wide dissemination network.
- Offer faculty development programming on the use of these fair testing guidelines in the creation of program testing and progression policies.
- Promote research on and development of fair testing practices.
- Lead in efforts to ensure that assessments of the costs and benefits of various testing schemas and testing policies consider the differential impact they have on minority and underrepresented groups in nursing education.
- Promote the development and validation of new approaches to assessment of the knowledge, skills, and competencies essential for quality nursing practice.
- Urge boards of nursing to support ethical and evidence-based fair testing practices to assess student learning in nursing education programs in ways that are themselves more inclusive of factors of import prior to NCLEX outcomes.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing* (5th ed.). American Educational Research Association.

Bowers, A. J., & Zhou, X. (2019). Receiver operating characteristic (ROC) area under the curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes. *Journal of Education for Students Placed at Risk* (JESPAR), 24(1), 20-46. https://doi.org/10.1080/10824669.2018.1523734

Brodersen, L. D., & Mills, A. C. (2014). A comparison of two nursing program exit exams that predict first-time NCLEX-RN outcome. *Computers, Informatics, Nursing: CIN*, 32(8), 404-412. https://doi.org/10.1097/CIN.0000000000000081

Helms, J. (2005). Stereotype threat might explain the black-white test-score difference. *American Psychologist*, 60(3), 269-270. http://doi.org/10.1037/0003-066X.60.3.269

Helms, J. (2006). Fairness is not validity or cultural bias in racial-group assessment: A quantitative perspective. *American Psychologist*, 61(8), 845-859.

Hunsicker, J., & Chitwood, T. (2018). High-stakes testing in nursing education: A review of the literature. *Nurse Educator*, 43(4), 183-186. https://doi.org/10.1097/nne.0000000000000475

National League for Nursing. (2011). *Annual Survey of Schools of Nursing 2011*. Author.

Odom-Maryon, T., Bailey, L. A., & Amiri, S. (2018). The influences of nursing school characteristics on NCLEX-RN® pass rates: A national study. *Journal of Nursing Regulation*, 9(3), 59-69. https://doi.org/10.1016/s2155-8256(18)30154-6

Plake, B. S., & Wise, L. L. (2014). What is the role and importance of the revised AERA, APA, NCME Standards for Educational and Psychological Testing? *Educational Measurement: Issues and Practice*, 33(4), 4-12. https://doi.org/10.1111/emip.12045

Randolph, P. K. (2017). Standardized testing practices: Effect on graduation and NCLEX® pass rates. *Journal of Professional Nursing*, 33(3), 224-228. https://doi.org/10.1016j.profnurs.2016.09.002

Smolkowski, K., & Cummings, K. D. (2015). Evaluation of diagnostic systems: The selection of students at risk of academic difficulties. *Assessment for Effective Intervention*, 41(1), 41–54. https://doi.org/10.1177/1534508415590386

Solórzano, R. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research*, 78(2), 260-329.

Spector, N., Silvestre, J., Alexander, M., Martin, B., Hooper, J. I., Squires, A., & Ojemeni, M. (2020). NCSBN regulatory guidelines and evidence-based quality indicators for nursing education programs. *Journal of Nursing Regulation*, 11(2, Supplement), S1-S64. https://doi.org/10.1016/S2155-8256(20)30075-2

Spurlock, D. R. (2006). Do no harm: Progression policies and high-stakes testing in nursing education. *Journal of Nursing Education*, 45(8), 297-302. http://doi.org/10.3928/01484834-20060801-04

Spurlock, D. (2013). The promise and peril of high-stakes tests in nursing education. *Journal of Nursing Regulation*, 4(1), 4–8. https://doi.org/10.1016/s2155-8256(15)30172-1

Spurlock, D. R., & Hunt, L. A. (2008). A study of the usefulness of the HESI Exit Exam in predicting NCLEX-RN failure. *Journal of Nursing Education*, 47(4), 157–166. https://doi.org/10.3928/01484834-20080401-07

Stout, W. (2002). Psychometrics: From practice to theory and back: 15 years of nonparametric multidimensional IRT, DIF/test equity, and skills diagnostic assessment. *Psychometrika*, 67(4), 485-518.

Wendt, A., & Kenny, L. (2007). Setting the passing standard for the National Council Licensure Examination for Registered Nurses. *Nurse Educator*, 32(3), 104-108. http://doi.org/10.1097/01.NNE.0000270234.99531.07

Zieky, M., & Perie, M. (2006). *A primer on setting cut scores on tests of educational achievement*. Princeton, NJ: Educational Testing Service, Inc. https://www.ets.org/research/policy_research_reports/publications/publication/2006/dbkw

STRATEGIC ACTION GROUP

Darrell Spurlock, Jr. PhD, RN, NEA-BC, ANEF, Chair
Widener University

Pamela Randolph, MS, RN, FRE
Arizona State Board of Nursing

JoAnne Banks, PhD, RN
Winston-Salem State University

Meagan Rogers, PhD, RN-BC, CPEN
University of Texas at Arlington

Linda Christensen, EdD, JD, MSN, RN
National League for Nursing

Barbara Patterson, PhD, RN, ANEF, FAAN, Staff Liaison
National League for Nursing

Kyle Johnson, PhD, RN, CHSE
Texas Tech University Health Sciences Center