

NLN Assessment Services Division

NLN NACE Series

Technical Manual



Table of Contents

Preface.....	3
The National League for Nursing	4
___ NLN Mission.....	4
___ Overview of the NLN	4
THE NLN NACE Exam Series.....	5
___ Purpose and Overview of the NLN NACE Exams	5
___ Description of the NACE Exams	5
Review of Test Specifications	8
___ Content Specifications.....	8
___ Statistical specifications	14
___ Selection of Item Writers	14
___ Item Construction	15
___ Review of Items.....	16
Technical Characteristics of the NACE Exam Series	17
___ Experimental Testing	17
___ Equating and Item Banking	17
___ Annual Review	18
___ Setting Benchmarks for NLN NACE Exams	18
Scoring NLN NACE Exams and Reporting Results.....	19
___ Criterion – Reference Results	19
___ NACE Score Reports	20
___ Quality Control of Scoring	21
Psychometric Quality	22
___ Validity Evidence.....	24
Other Information Regarding NLN Test Administration	26
___ Limitations of Exams	26
___ Test Policy.....	26
___ Test Administration	26
___ Retake Policy	27
___ Fair Testing and Ethics Statement	27
___ Special Accommodations	28

References.....29

Table of Tables

Table 1 Foundations of Nursing Topics.....9
Table 2 Nursing Care of the Adult Client Topics.....10
Table 3 Care of the Child Topics.....11
Table 4 Care of the Childbearing Family Topics.....12
Table 5 Care of the Client with Mental Disorder Topics.....13
Table 6 NACE Exams Analyses.....23

Preface

This manual contains technical information about the NLN Nursing Accelerated Challenge Exams NACE revised in 2019-2020. The purpose of this manual is to provide the technical characteristics of NACE exams as they were intended. The National League for Nursing (NLN) regularly conducts an analysis of the need and use for the NACE exams. These findings are evaluated to ensure the exam series continues to be rigorous.

The content of this manual responds to requirements of the testing industry as stated by the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), and the Fair Testing Guidelines for Nursing Education (National League for Nursing, 2012).

The National League for Nursing

NLN Mission

The National League for Nursing promotes excellence in nursing education to build a strong and diverse nursing workforce to advance the health of our nation and the global community.

Overview of the NLN

Dedicated to excellence in nursing, the National League for Nursing is the premier organization for nurse faculty and leaders in nursing education. The NLN offers professional development, networking opportunities, testing services, nursing research grants and public policy initiatives to its 40,000 individual and 1,200 institutional members. NLN members represent nursing education programs across the spectrum of higher education, health care organizations and agencies.

Founded in 1893 as the American Society of Superintendents of Training Schools for Nurses, the NLN was the first nursing organization in the United States. Headquartered in Washington, D.C., the NLN is led by a board of governors elected by membership for three-year terms. The volunteer board chair works with the NLN's president and chief executive officer.

THE NLN NACE Exam Series

Purpose and Overview of the NLN NACE Exams

The National League for Nursing (NLN) offers the Nursing Acceleration Challenge Exam (NACE) to facilitate educational mobility by providing educators with an instrument for diagnostic evaluation of the nursing knowledge of individuals seeking advanced placement in a nursing program. The NACE is designed to evaluate previous learning and experience. It is intended to assist faculty to establish credit and to make placement decisions about students already selected for admission to the program. Information from the NACE can be used as one component of the total evaluation of each individual.

Description of the NACE Exams

The NACE includes five content areas. Programs may elect to have students take any one of these exams.

Foundations of Nursing: Includes content related to the basic physiological and psychosocial needs of individuals. (200 Items / 4 hours administration)

Care of the Adult Client: Includes content related to drug administration, nutrition and diet therapies, and nursing interventions for care of individuals. (200 Items / 4 hours administration)

Care of the Child: Includes content related to nursing care of the infant, the toddler and preschooler, and the school-aged child and adolescent. (100 Items / 2 hours administration)

Care of the Client During Childbearing: Includes content related to nursing care during the antepartal, intrapartal, postpartal and neonatal periods. (100 Items / 2 hours administration)

Care of the Client with Mental Disorder: Includes content related to nursing care of individuals who have mental disorders. (100 Items / 2 hours administration)

The test items constitute a representative sample of health problems and conditions commonly encountered in nursing. The questions relate to the promotion, maintenance, and restoration of health. They emphasize normal findings, as well as deviations from normal treatment modalities, which include drugs and nutrition, and nursing interventions. The division of the content reflects courses most commonly found in nursing schools.

The NACE is designed to measure objectives derived from a thorough analysis of nursing curricula and defined in consultation with a variety of nurse educators at this level. These objectives are:

- (A) The nurse will apply and demonstrate knowledge of the individual client's state of health including biological, psychological, and sociological status.
- (B) The nurse will apply and demonstrate knowledge of nursing measures appropriate for the care of clients.

(C) The nurse will apply and demonstrate knowledge of drugs, including intended and unintended effects, used in the treatment of clients.

(D) The nurse will apply and demonstrate knowledge of normal nutrition and therapeutic diets used in the treatment of clients.

NACE items are written in the framework of the nursing process. For purposes of the NACE, the steps of the nursing process are defined as follows:

- **Assessment:** Includes collecting physical, emotional, and cultural data; recognizing signs, symptoms, and behaviors.
- **Analysis:** Includes interpreting data, identifying clients' needs, and making nursing diagnoses.
- **Planning:** Includes setting goals, selecting measures for care, and identifying correct techniques for giving care.
- **Implementation:** Includes responding appropriately to clients needing information and/or teaching.
- **Evaluation:** Includes recognizing intended and unintended effects of measures and treatments, determining evidence of the client's compliance with the treatment plan and determining evidence of the client's response to measures and treatment.

Faculty within a nursing program are responsible for determining the appropriateness of one or more of the NACE tests for the program. The content outline and sample questions included in this manual can be useful in making this decision.

Review of Test Specifications

Two types of test specifications are used in developing the NACE exam content specifications and statistical specifications.

Content Specifications

Content specifications for the NACE exams were developed through the curricular analysis discussed above. While the purpose and objectives of the exams remain the same from year to year, so scale scores are comparable, the content areas of test items used are reviewed regularly. Test Construction Committees are convened to review the item banks to verify content accuracy and match the content of the tests to content specifications. At these committee meetings, the characteristics of the items that fulfill the content specifications are also reviewed. While the general content of the test remains constant, particular kinds of items in a specification category may change. Each NACE exam administration is comprised of items randomly selected from the related item banks. The related topic areas for each exam title are provided in the tables below.

Table 1
Foundations of Nursing Topics

Safe & Effective Care Environment	Health Promotion & Maintenance
Home health Chronic illness Palliative care/Hospice Documentation Pre- and post-op care Infectious disease Sterile technique Safe ambulation	Respiration/circulation Vital signs Comfort Communication -Verbal -Non-verbal Moral/ethical/legal issues Nutrition Exercise Stress management
Psychosocial Integrity	Physiological Integrity
Addiction/recovery Loss and death Fractured lives Meditation/mindfulness Transgender care Self-actualization Security Self-esteem Spiritual & cultural issues	Basic health needs Hygiene/incontinence Elimination Medication administration Fluids and electrolytes IV therapy Body alignment Sensory changes in the elderly Obesity

Table 2
Nursing Care of the Adult Client Topics

Safe & Effective Care Environment	Health Promotion & Maintenance
Metabolic regulation	Management of chronic health conditions
Transitions of care	Screening for prevention
Care settings	Provision of support services
Environmental accommodations	Accommodations for independence
Psychosocial Integrity	Physiological Integrity
Neurological disorders	Autoimmune disorders
Developmental needs of aging adults	Blood dyscrasias
Integumentary conditions	Cardiac & vascular conditions
Family and caregiver support	Metabolic regulation
	Neurological disorders
	Gastrointestinal conditions
	Genitourinary conditions
	Integumentary conditions

Table 3
Care of the Child Topics

Safe & Effective Care Environment	Health Promotion & Maintenance
Environmental accommodations Interventions for unsafe environments Hospitalization and chronic illness	Routine health screening and immunizations Family and caregiver support Nutrition needs Normal growth milestones
Psychosocial Integrity	Physiological Integrity
Promotion of growth and development Mental health Developmental delay issues Educational concerns	Autoimmune disorders Blood dyscrasias Cardiac and vascular conditions Metabolic regulation Neurological disorders Gastrointestinal conditions Genitourinary conditions Integumentary/allergy conditions Musculoskeletal conditions Pulmonary conditions Sensory alterations and accommodations

Table 4
Care of the Childbearing Family Topics

Safe & Effective Care Environment	Health Promotion & Maintenance
Inpatient care Preparation for labor & delivery Appropriate ways to deal with discomforts of pregnancy Antepartum testing Labor and delivery complications Potential complications in postpartum period Potential complications of the neonatal period	Prenatal Care Communication Moral/ethical/legal issues Nutrition Exercise Stress management Development tasks of becoming parents / Pre-conception care Teaching/education provision
Psychosocial Integrity	Physiological Integrity
Substance abusing clients and their infants Meditation/mindfulness Transgender/alternative family care Self-actualization Security Self-esteem Spiritual and cultural issues Anxiety	Basic health needs Mental health changes postpartum Obesity Nutritional needs *During pregnancy *During lactation High risk pregnancies

Table 5
Care of the Client with Mental Disorder Topics

Safe & Effective Care Environment	Health Promotion & Maintenance
Chronic mental illness management Documentation/compliance Homelessness Inpatient care Caregiver role strain	Communication -Verbal -Non-verbal Moral/ethical/legal issues Nutrition Exercise Stress management
Psychosocial Integrity	Physiological Integrity
Addiction/recovery Suicide Fractured lives Meditation/mindfulness Transgender care Self-actualization Security Self-esteem Spiritual & cultural issues Adjustment disorders Anxiety	Basic health needs Mental health changes in the elderly Obesity Neurodevelopmental disorders *Dementia *ADHD *Autism Feeding and eating disorders Psychophysiological mental disorders *Schizophrenia *Depression *Bipolar *Obsessive-compulsive

Statistical Specifications

The difficulty of an item on an NLN exam is measured by the proportion of applicants who answer the test item correctly. At NLN, this proportion-correct value for each item is transformed to a difficulty index on a standard scale (the delta scale). An index of item discrimination is also calculated. The discrimination index for an item is the point biserial correlation between applicant's performance on an individual item and their overall performance on the measure. It is expected that, as a group, applicants who perform well on the total test will perform well on an individual item.

A range of item difficulties is required in each test form in order to differentiate among examinees at different levels of abilities measured by exam subcategories. Pretesting items provides appropriate data regarding difficulty levels, discrimination indices and estimates of reliability. Before the operational test form is published, an additional review is conducted by the test development team, including subject matter experts and a psychometrician. The final operational test banks are constructed and verified through an item analysis. Ideally, an item is considered "good" if the discrimination index is above .30; "fair" if it is between .10 and .30; and "poor" if it is below .10. NLN aims to have an average difficulty of 0.60 with a range from 0.2 to 0.89 and an average item discrimination index of about 0.40.

Selection of Item Writers

Each year, NLN contracts with item writers to construct items for NLN exams. With the exception of the item writers for pre-admission exam, all item writers are actively engaged or recently retired nurse educators in nursing programs from a variety of

institutions. NLN makes every attempt to include item writers who represent the diversity of the United States' population with respect to ethnic background, gender and geographic location. Before being asked to write items for NLN tests, potential item writers are required to apply to the NLN. Then, the item writer receives an NLN item writer's guide. The guides include examples of items and provide item writers with the test specifications and NLN's requirements for content and style. Included are specifications for fair portrayal of all groups of individuals, avoidance of subject matter that may be unfamiliar to members of certain groups within society, and nonsexist use of language. Each application submitted by a potential item writer is evaluated by the NLN Development staff. A decision concerning whether to contract with the item writer is made on the basis of that evaluation. Each item writer under contract is given an assignment to produce a small number of items. The small size of the assignment ensures production of a diversity of material and maintenance of the security of the testing program, since any item writer will know only a small proportion of the items produced. Item writers work closely with NLN content specialists, who assist them in producing items of high quality that meet the test specifications.

Item Construction

Item writers must create items that are written in the NCLEX style and are psychometrically sound. A number of items must be constructed because, even with good writers, many items fail to meet NLN's standards. Each item writer submits a set of items in a given content area.

Review of Items

After items are accepted, there is a rigorous process to ensure items are approved by the Test Construction Committee. They are checked to meet NLN's specifications for accuracy, item classification, item format and language. The items are reviewed several times by NLN staff to ensure they meet all NLN's standards. Items are then submitted to content and fairness committees for external reviews before the pretest administration of the items. The content committee reviews the items for content accuracy, educational importance and NCLEX appropriateness. The fairness review committee consists of experts in diverse educational areas who represent both genders and a variety of racial and ethnic backgrounds. The fairness committee reviews the items to help ensure fairness to all examinees. Any comments by the content consultants are discussed in a meeting with all the content consultants and NLN staff, and appropriate changes are made to the item(s). All fairness consultants' comments are reviewed and discussed, and appropriate changes are made to the item(s).

Technical Characteristics of the NACE Exam Series

Experimental Testing

Items that are judged to be acceptable in the review process are added to tryout topics for pretesting on sample groups from the national examinee population. These sample groups are carefully selected to be representative of the total examinee population. Each sample is administered tryout topics for an exam title. The time limits for the tryout items permit the majority of students to respond to all items. Item analyses are then performed on the tryout topics. Biserial and point-biserial correlation coefficients between each item score (correct/incorrect) and the total score on the corresponding test of the standard exam are also computed. Item analyses serve to identify statistically effective test items. Items that are either too difficult or too easy are eliminated or revised for future item experimental testing. The biserial and point-biserial correlation coefficients, as well as the differences between proportions of students answering the item correctly in each of the three groups, are used as indices of the discriminating power of the tryout items. Each item is reviewed after the item analysis. NLN staff members scrutinize items flagged for statistical reasons to identify further review. The review process also provides feedback that helps decrease the incidence of poor items in the future.

Equating and Item Banking

Each test administration is equated to a standard reference administration so that the reported scores across all forms and years are comparable. The method that is

employed to adjust scores on administrations that differ in difficulty requires that the administrations share some items in common. The equating or anchor items are then used to place scores on different exam administrations on the same score scale. This way, applicants with the same ability are expected to receive the same score regardless of which items are generated for test administration.

Annual Review

Every 12 months, the NACE exams are psychometrically reviewed item analysis results are reviewed for any anomalies - changes in item difficulty and discrimination indices between experimental testing and operational administrations. Examinees may challenge any items that they feel are questionable. Once a challenge to an item is raised and reported, the item is reviewed by content specialists in the content area assessed by the item. In the event that a problem is found, actions are taken to eliminate or minimize the influence of the problem item as necessary. In all cases, the person who challenges an item is sent a letter indicating results of the review. Every three years, a more extensive review is completed by the development team to review and revise the test blueprint in addition to the psychometric analysis.

Setting Benchmarks for NLN NACE Exams

Each program determines the credit to be given for acceptable performance on a NACE test. NLN is not involved in this procedure, nor does NLN establish a "passing" or "failing" mark on these tests. Each nursing program is encouraged to set its own performance standards by matching a NACE test with a specific course or course

sequence and determining acceptable test performance commensurate with acceptable performance of students already enrolled in the program. Our test development staff will aid the program if requested.

Scoring NLN NACE Exams and Reporting Results

Criterion – Reference Results

Assessment reports will be electronically available in the school and student accounts within 24 hours of testing, unless otherwise stated. The NACE uses a criterion-referenced approach. An overall exam score and diagnostic sub-scores are reported as percentages of questions correct. The average percentages correct for all nursing student examinees are provided as well. These are based on content areas in the individual tests. The information is intended to supplement faculty evaluation of students and students' own self-assessment of strengths and weakness.

Group average percent-correct scores do not in any way constitute recommendations for placement standards or benchmark scores, which must be established by the individual institutions. Indeed, the average of the national sample of students lies well above the minimum level required for passing by most institutions.

Test scores are the property of NLN and are retained on file in electronic form for a period of three years after the date of examination. Duplicate copies of score reports are available upon written request by submitting the appropriate request form.

NACE Score Reports

There are different reports available for the NACE exam. The first report type for examinee and the school administrator is the coaching report. The coaching report shown below in Figure 1, provides the student and the program with a comparison of the examinee performance versus the average performance for each subcategory and overall score.

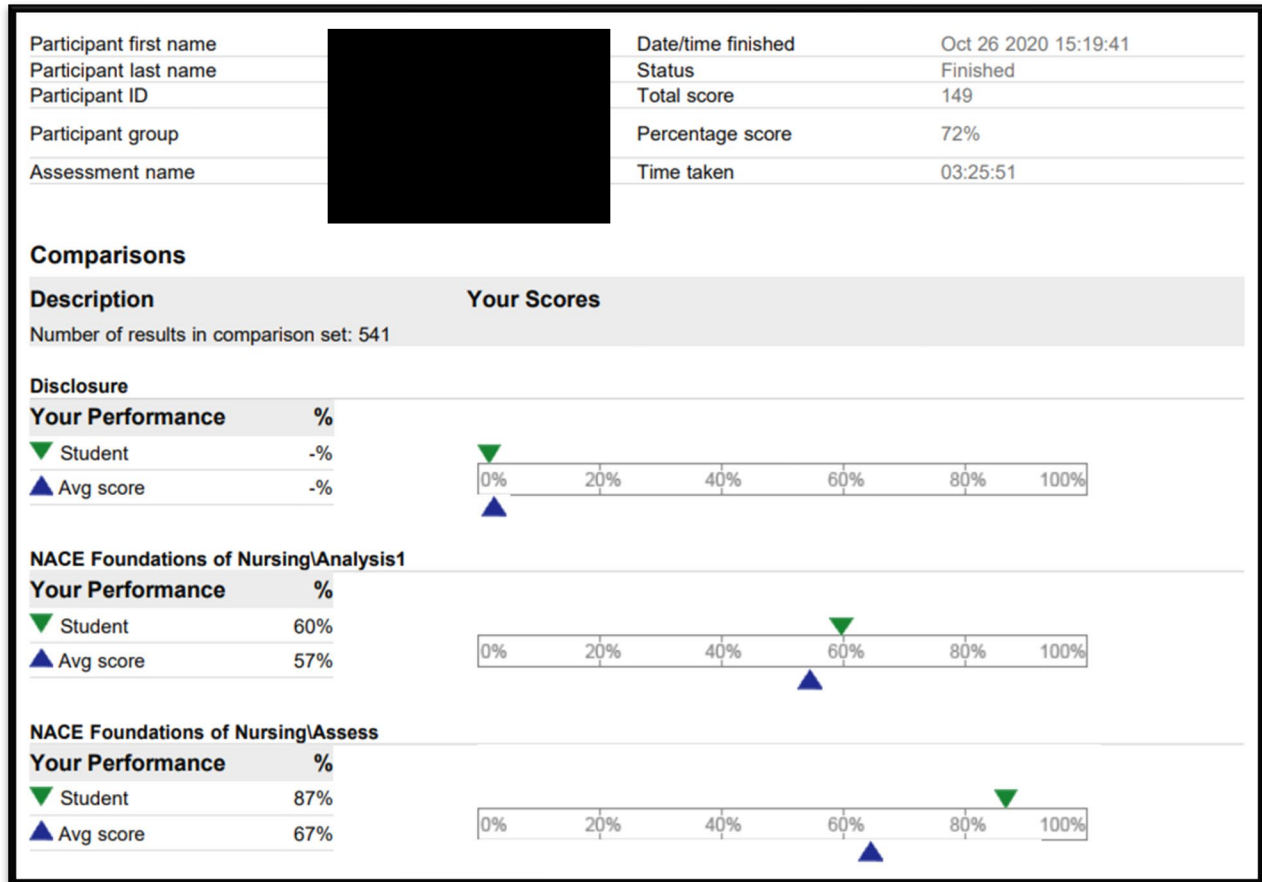


Figure 1 NLN NACE Coaching Report

A	B	C	D	E	F	G	H
Score ▾	% ▾	Completed Date ▾	Analysis % ▾	Assess % ▾	Evaluating % ▾	Implementation % ▾	Planning % ▾
150	75	12/3/2020 4:16	63	73	80	75	83
127	63	12/3/2020 21:12	63	63	63	68	53
116	58	12/14/2020 17:40	47	53	73	56	63
130	65	12/21/2020 21:47	40	77	70	68	67

Figure 2 NLN NACE Score List Report

The Score List report shown above in Figure 2 provides an aggregate list of scores within a selected date range. Overall and/or topic scores may be viewed.

Quality Control of Scoring

Accuracy of the scores is verified through the use of quality control procedures that are followed through every stage of the test-scoring process. There are established procedures for the quality control of the scores that are reported to examinees and institutions. A sample of answer sheets is randomly selected and independently processed at key points in the scoring and reporting operations. Although the NLN is committed to reporting scores in a timely manner, score accuracy is the first priority, and scores will not be released until the quality control process is complete and the accuracy of the scores has been verified. The NLN routinely follows extensive review and quality control procedures to detect and avoid flawed questions and scoring errors. However, occasionally, an error may be detected after scores have been reported. In the event of such an error, the program recalculates the scores and reports revised scores for those students whose scores change. Revised scores reported during the year the test was given are reported directly to nursing schools as well as to students.

Psychometric Quality

In the educational measurement field, two primary criteria are used to define high-quality information validity and reliability (Messick, 1989). Indeed, these two criteria constitute the first two chapters of the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999). Validity is concerned with the appropriateness, meaningfulness and accuracy with which inferences can be made about a measured characteristic; reliability deals with the consistency of measures under different conditions (Messick, 1989; AERA, APA, & NCME, 2014). This section presents the data collected from the 2019 sample groups and estimates of the reliability of NACE scores (Cronbach's Alpha and Standard Error of Measurement).

Table 6
NACE Exams Analyses (2020)

	Foundations of Nursing		Care of the Adult Client		Care of the Child		Care of Childbearing Family		Care of Client w/ Mental Disorder	
	# of Items	%	# of Items	%	# of Items	%	# of Items	%	# of Items	%
Mean Overall Score	200	72	200	67	100	65	100	64	100	63
SD	14.8				8.2		9.3		10.1	
Assessment	30	80	50	63	20	62	20	63	20	63
Analysis	30	50	30	68	20	66	20	68	20	54
Planning	30	69	30	59	20	63	20	59	20	60
Implementation	80	77	50	66	20	64	20	66	20	68
Evaluation	30	73	40	72	20	70	20	72	20	61
Reliability	.6		.6		.5		.6		.7	
SEM	8.93		11.6		6.23		5.84		5.56	

Mean and **standard deviation (SD)** are based on the uncorrected raw score, i.e., the number of items answered correctly.

The **reliability coefficient** is an estimate of the consistency with which the trait under consideration is being measured. The reliability estimate reported here is Cronbach's Alpha, using raw score data. The range of the coefficient is between 0 and +1. This estimate is based on inter-item consistency, requiring only one administration of the test.

Standard error of measurement (SEM) is reported to help assess the accuracy of measurement of an individual's score. The SEM estimates the variation in a score one would expect to see if an individual were tested repeatedly. In practice, it is estimated

(with 68 percent confidence) that an individual's "true" score will fall within ± 1 SEM of the observed, or actual, score on the test. Therefore, small differences are not likely to indicate significant differences among the performances of students. (To obtain the value of the SEM, we employ values of the standard deviation and reliability that have greater precision than shown in the above table.)

Validity Evidence

According to the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), "validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (p. 9). Arguments for the validity of an intended inference made from a test may contain logical, empirical, and theoretical components. A distinct validity argument is needed for each intended use of a test. The potential interpretations and uses of NACE scores are numerous and diverse, and each needs to be justified by a validity argument.

Detailed test specifications have been developed to ensure that the test content is representative of current RN nursing programs. All test items are reviewed to ensure that they match these specifications. Hence, there is an ongoing assessment of the content validity of the tests during the development process.

The standardization of the NACE exams is also important to their proper use as measures of educational achievement. The courses students take in nursing school and the grades they earn are also measures of educational achievement, but these variables are not standardized measures. They cannot be standardized because course content varies considerably among schools of nursing, and grading policies certainly

vary among instructors. Therefore, while all courses taken and grades earned are measures of educational achievement, their interpretation should properly consider differences in nursing school curricula and grading policies.

Setting a benchmark based on average NACE scores may be less valid given the differences discussed above.

To set a benchmark, programs should administer the NACE exam to a sample set of students who successfully completed the course for which the NACE exam is being challenged. The scores may then be compared with students challenging the course to make a benchmark decision.

Other Information Regarding NLN Test Administration

Limitations of Exams

Like any exam, the Nursing Acceleration Challenge Exam (NACE) is not a perfect measurement of achievement. Even “comprehensive” tests can only sample knowledge in a given field. Because of this, questions on a specific test may be more favorable to some examinees than to others. Errors in measurement may also be introduced by examinees. Individuals who are ill or extremely tired or nervous may not perform to the best of their ability. The NACE is not designed to point out health or personal factors that impinge upon educational success. These limitations should not be forgotten by those reviewing students’ performance whether for challenge or for other purposes.

Test Policy

The NLN Test Policy was developed to ensure the security of our tests and to protect their integrity. All NLN Tests are copyright-protected. No portion of any test may be reproduced without the permission of the NLN.

Test Administration

The NACE is administered by faculty at individual schools of nursing. Any official representative of the school or program may be designated as an administrator or proctor and authorized to administer the tests. Additionally, NLN authorizes the administration of the NACE exam via our contracted vendor for remote, one-on-one proctoring. Contact the NLN Assessment Services Division for additional information.

Retake Policy

The NLN recommends that students not be allowed to re-test for a period of six months or longer. However, it is up to the institution to set the re-test policy. The NLN does not regulate the re-taking of an exam and will report ALL scores attained by the examinee.

Fair Testing and Ethics Statement

We believe most examinees manage themselves with integrity and are disturbed when they observe others cheating. Examinees are expected to maintain the highest standards of conduct. Using or attempting to use unauthorized assistance, material or study aids during an examination, will result in a disqualification of the exam and future NLN exams. Acts of misconduct include but not limited to:

- Copying from others.
- Providing or receiving information about all or part of an exam, including answers (e.g., telling someone or publishing what was on your exam or being told this information).
- Having or using a “cheat sheet” (i.e., a piece of paper with answers, formulas, information or notes) that is not specifically authorized.
- Having or using a communication device such as a cellphone, tablet or alike to send or obtain unauthorized information.
- Taking an exam for another student or permitting someone else to take a test for you.
- Asking another individual to give you improper assistance, including offering money or other benefits.

- Asking for or accepting money or any other benefit in return for giving another individual improper assistance.
- Altering an assessment report and resubmitting it.
- Gaining or providing unauthorized access to examination materials.

Note: Simply having possession during an exam of any prohibited or unauthorized information or device, whether or not it is actually used, is an act of dishonesty and will be dealt with as such.

Special Accommodations

In accordance with the Americans with Disabilities Act of 1990, the NLN has established a Testing Accommodations Policy to eliminate discrimination against individuals with disabilities and to adhere to the standards and protections established on behalf of those individuals. The policy is applicable to all persons requesting special testing accommodations from the NLN. Requests for special testing accommodations must be made through each college's office for students with disabilities. An application for special testing accommodations containing instructions and related pertinent information must be submitted for review with a minimum of five business days before testing.

References

- American Educational Research Association [AERA], American Psychological Association [APA], & National Council for Measurement in Education [NCME]. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- American Educational Research Association [AERA], American Psychological Association [APA], & National Council for Measurement in Education [NCME]. (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Clauser, B., & Mazor, K. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31-44.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Harcourt Brace Jovanovich.
- Hambleton, R. (1980). Review methods for criterion-referenced test items. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.
- Jaeger, R. (1991). Selection of judges for standard-setting. *Educational Measurement: Issues and Practice*, 10(2), 3-14.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4), 719-747.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (Vol. 3, p. 13-103). American Council on Education/Macmillan.
- Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Office of Federal Contract Compliance Programs [OFCCP], Equal Employment Opportunity Commission [EEOC], & Department of Labor [DOL]. (1978). *Uniform guidelines on employee selection procedures* (41. CFR Part 603). Washington, DC: Equal Employment Opportunity Commission.
- Schmeiser, C., & Welch, C. (2006). Test development. In *Educational measurement* (4th ed., p. 307-353). Westport, CT: American Council on Education/Praeger.