PREDICTING NCLEX SUCCESS WITH THE HESI EXIT EXAM:

RESULTS FROM FOUR YEARS OF STUDY

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

COLLEGE OF NURSING

BY

AINSLIE T. NIBERT, M.S.N.

DENTON, TEXAS

MAY 2003

TEXAS WOMAN'S UNIVERSITY
DENTON, TEXAS
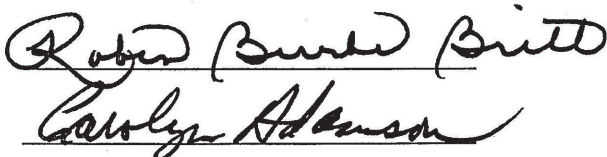
_____March 12, 2003_____
Date

To the Dean of The Graduate School:

I am submitting herewith a dissertation written by Ainslie T. Nibert entitled
"Predicting NCLEX Success with the HESI Exit Exam: Results from Four Years of
Study." I have examined this dissertation for form and content and recommend that it
be accepted in partial fulfillment of the requirements for the degree of Doctor of
Philosophy with a major in Nursing.

_____
Anne Young, Major Professor

We have read this dissertation and
recommend its acceptance:

Accepted:

_____
Dean of The Graduate School

ACKNOWLEDGEMENTS

ABSTRACT

AINSLIE NIBERT

PREDICTING NCLEX SUCCESS WITH THE HESI EXIT EXAM:

RESULTS FROM FOUR YEARS OF STUDY

MAY 2003

The fourth annual validity study of the Health Education Systems, Inc. (HESI)

Exit Exam (E^2) was designed to examine not only the accuracy of the E^2 in predicting

NCLEX success, but also the degree of risk for failure of the licensure exam associated

with specific scoring intervals. A descriptive, correlational design was used to examine

the data provided by schools of nursing regarding students' NCLEX outcomes in the

academic year 1999-2000. As in the three previous studies, the E^2 was found to be a

highly accurate predictor of NCLEX success (98.46%). Each scoring interval was

significantly different from each of the other scoring intervals ($P = .001$). In fact, for the

combined group of RN and PN students, the percentage of students who failed the

NCLEX more than doubled with each successively lower scoring interval. Of the 2,059

RN students who scored in the A/B category, 35 (1.70%) failed the licensing exam; of the

1,014 students who scored in the C category, 60 (5.92%) failed; of the 980 students who

scored in the D category, 106 (10.82%) failed, of the 1,324 students scoring in the E/F

category, 314 (23.72%) failed, and of the 526 students scoring in the G/H category, 264

(50.19%) failed. These findings provide the information faculties need to make evidence-

based decisions regarding students' risk for NCLEX failure. Additionally, frequency data were obtained from this survey regarding the use of the E^2 as a benchmark for progression and remediation, and these findings may also be useful to faculties that are considering establishment of such programs.

This study investigated use of the HESI Exit Exam (E^2) as a benchmark for progression. Data obtained from 158 schools of nursing indicated that: (1) faculties are increasingly adopting policies that specify E^2 scores as benchmarks for progression; (2) a HESI score of 85 is the most frequently used benchmark; and (3) completion of a remediation course followed by mandatory re-testing is the strategy most frequently prescribed for those who fail to achieve designated benchmarks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

While there is an ongoing nursing shortage in the United States, one of unprecedented scope will grip the nation by 2020, fueled by the retirement of almost half of the current "baby boomer" generation of nurses who are being replaced by fewer numbers of the next generation (Buerhaus, Staiger, & Auerbach, 2000; United States Census Bureau, 2000). Currently, decreasing numbers of candidates are taking nursing licensure examinations. In 1999, there were 113,247 candidates tested for the National Council Licensure Examination for Registered Nurses (NCLEX-RN) compared to 116,713 during 1998, which represents a decrease of nearly 3%. An even greater decrease occurred in the number of practical nursing (PN) candidates, with 47,592 PN candidates testing for the NCLEX-PN in 1999, compared to 50,230 in 1998, representing a decrease of approximately 5.3% (National Council of State Boards of Nursing, 2001).

Not only are the numbers of nursing students decreasing, but also pass rates on nursing licensure exams. From 1995 through 2000, the annual pass rates for both the NCLEX-RN and National Council Licensure Exam for Practical Nurses (NCLEX-PN) have decreased in each successive year. For the registered nurse (RN) licensure exam, the annual pass rate decreased from 90.2% in 1995 to 83.8% in 2000, and for the PN licensure exam, the annual pass rate decreased from 90.8% in 1995 to 85.0% in 2000 (National Council of State Boards of Nursing, 2001). Recruitment of a more culturally

diverse candidate pool has been described as a possible solution to the documented nursing shortage (American Association of Colleges of Nursing, 2000a, 2000b; Grossman et al., 1998). However, previous studies indicate that ethnic minorities and foreign-born nursing students experience higher attrition rates and higher National Council Licensure Examination (NCLEX) failure rates than their non-minority, English-speaking counterparts (Arathuzik & Aber, 1998; Endres, 1997; Fearing, 1997; Frierson, Malone, & Shelton, 1993).

In response to these declining pass rates, nurse educators have attempted to identify students who are at risk for NCLEX failure as early as possible so that remediation can be promptly initiated. Nursing faculties have increasingly turned to the Health Education Systems, Inc. (HESI) Exit Exam (E^2) as a remediation guide, and more recently, as a benchmark for progression. HESI, founded by a nurse entrepreneur who has published nursing examinations for over a decade using a critical thinking approach, presently accommodates approximately 400 schools of nursing with enrollments of more than 10,000 students annually with computerized custom and standardized examinations. Three previous studies validated the Health Education Systems, Inc. (HESI) Exit Exam (E^2), a comprehensive computerized nursing examination, as a predictor of NCLEX success (Lauchner, Newman, & Britt, 1999; Newman, Britt, & Lauchner, 2000; Nibert & Young, 2001) as well as NCLEX failure (Hanks & Lauchner, 1999).

A new and growing trend among nursing faculties is to set benchmarks near the end of the curriculum that contain progression to graduation requirements based on objective evidence of nursing competency. These policies are designed to identify

2

students in need of remediation prior to graduation and NCLEX candidacy so that remediation can be initiated and NCLEX failure can be avoided. Morrison, Free, & Newman (2002) found that progression policies were highly effective in increasing pass rates. However, to date, no research had been conducted regarding the degree of risk for licensure failure associated with specific HESI scores. Previous studies examined only high-scoring and low-scoring E^2 students' NCLEX success, and these two scoring categories were not specific enough to assist faculty in defining E^2 benchmarks (Lauchner et al., 1999; Newman et al., 2000; Nibert & Young, 2001). Consequently, faculties' decisions regarding minimally acceptable E^2 scores have been based on professional judgment rather than evidence-based research.

Problem of Study

This study was designed to establish the predictive validity of the HESI E^2 with regard to success on the NCLEX-RN and NCLEX-PN through an analysis of the degree of risk associated with various E^2 scoring intervals and to investigate the uses of the E^2 as a benchmark for progression and guide for remediation in nursing curricula. Specifically, the study's purposes included: determining the predictive accuracy of the E^2 for students in all types of nursing programs; identifying criteria used by educators to determine students' readiness for graduation as stated in policies of schools administering the E^2; and describing remediation strategies designed to assist students who fail to achieve minimal E^2 benchmarks.

Rationale for the Study

NCLEX failure not only contributes to the nursing shortage by delaying new graduates' entrance into the work force, but such failures also have personal and financial consequences for candidates, nursing faculties and administrators, and prospective employers. Unsuccessful NCLEX candidates suffer loss of potential wages that might have been earned if they were licensed nurses. In addition to the financial consequences of licensure failure for unsuccessful NCLEX candidates, Vance and Davidhizar (1997) reported that failure results in an even greater emotional loss, characterized by feelings of inadequacy and grief.

NCLEX failures also have financial consequences for health care employers because hiring and orienting new graduates is a costly institutional expenditure. Messmer, Abelleira, and Erb (1995) estimated nursing orientation costs to be between $20,000 and $50,000 per person. Licensure failure negates any benefit of such expenditures because candidates failing the NCLEX cannot assume the licensed nursing positions for which they were hired and oriented to fill.

NCLEX pass rates affect a school's reputation, thereby having consequences for nursing faculties and administrators. The public's view of a school can affect the ability to recruit new students into nursing programs. The two national nursing accrediting agencies, the National League for Nursing Accrediting Commission (NLNAC) and the Commission on Collegiate Nursing Education (CCNE), as well as the approval standards of most states' Boards of Nurse Examiners (BNE), use pass rate data as benchmarks for program effectiveness (American Association of Colleges of Nursing, 1999; National

4

League for Nursing Accrediting Commission, 1999a, 1999b). A consistent pattern of low NCLEX pass rates can potentially place a nursing program's accreditation or state approval at risk, which may ultimately result in closure of the program.

Early identification of academically at-risk students in a nursing program coupled with early remediation to address their weaknesses in nursing content are the keys to averting NCLEX failures. While evidence exists indicating that remediation positively affects NCLEX success in students with known academic deficits, few authors have described specific benchmarks that identify students who are at risk of NCLEX failure and in need of remediation.

The findings of this study may be useful to nursing educators who are charged with making evidence-based decisions regarding the use of E^2 scores as benchmarks for progression and remediation in their programs. Ultimately, the application of such benchmarks, as identified through this study, may reduce the likelihood of NCLEX failure for future candidates. In the midst of declining student enrollment in the nation's schools of nursing, declining NCLEX pass rates, and a shrinking nursing workforce, increasing first-time NCLEX candidates' probabilities of success on the licensure exams is a top priority for nursing faculties. Based on data obtained from three previous studies, the E^2 has been determined to be highly accurate in predicting NCLEX success. NCLEX success remains the key to entry into the nursing workforce, and thus remains a significant part of the solution proposed to ameliorate the current nursing shortage.

Theoretical Framework

Establishing the predictive validity of the E^2 regarding outcomes on NCLEX is accomplished by applying concepts from both classical measurement theory and critical thinking theory. Figure 1 describes the model for test construction based upon classical test theory and critical thinking theory. The creation, administration, and interpretation of any test are accomplished through educational and psychological measurement processes. Crocker and Algina, (1986) stated that measurement of psychological attributes occurs when quantitative values are assigned to the sample of behaviors obtained by administering a test. By observing and classifying similar behaviors, the test designer is able to draw inferences about the psychological constructs that contribute to the make-up of the test taker. These authors further explain that, through the use of these constructs, the test designer can ultimately identify possible relationships between psychological constructs, or between constructs and practical consequences (Crocker & Algina, 1986). Thus, it is possible through empirical study to predict or control certain patterns of behavior. However, in order to effectively predict behaviors, such as entry-level performance of a registered nurse, the test designer must first quantify the observations that represent the constructs defining these behaviors. The designers of the E^2, a test that simulates the NCLEX, employed classical test theory in creating this measurement instrument, and revise it continuously based on reliability and validity indices, with the goal of predicting entry-level performance of registered nurses.

Sax (1997a) defined tests as series of tasks that are used to gather systematic observations that are presumed to be representative of educational or psychological

NCSBN Constructs	Clinical Nursing Knowledge of Test Item Writers	Validity
NCLEX test blueprints derived from job analysis studies	Nursing course objectives from syllabi	Assess for evidence of content, construct, and criterion-related validity as defined by classical test theory

Test Creation Phase

HESI Method for Writing Critical Thinking Test Items

✓ Include rationale for each test item

✓ Write questions at the application or above cognitive level

✓ Require multilogical thinking to answer questions

✓ Require a high level of discrimination to choose from among plausible alternatives

Test Evaluation Phase

Assumptions of Classical Test Theory

Measurement of Educational Attributes is Useful in Predicting Performance

True Score = Observed Score + Error

Bloom's Taxonomy (1956)

Paul's Critical Thinking Theory (1993)

Reliability
Estimate with KR-20s

Reliability
Item Analysis using difficulty index and point biserial correlations

Figure 1.   Conceptual Framework: Model for HESI Examination Construction

attributes. He further defined standardized tests as those constructed by test specialists who work with curriculum experts to allow comparison with an external representative student group, known as the norm group. The E^2 is a comprehensive, computerized, criterion-referenced test designed to be administered during the last semester or quarter of the curriculum that also allows for comparison with external norm groups and can therefore be classified as a standardized test within the content domain of nursing.

Crocker and Algina (1986) and Sax (1997b) identified the basic elements of a test constructed within the framework of classical test theory as reflective of a student's true (universe) score, obtained score, and error of measurement, or error score obtained from a sample of the student's behaviors and attributes detected through the administration of the test. The behaviors and attributes sampled on a single test are chosen by the test designer from the universe of test items that measure the universe of nursing concepts, known as the domain of professional nursing. The relationship between the scores obtained from a single test administration can be described by this formula: the obtained score equals the true score plus the error score. True, or universal, scores are hypothetical values that would represent the individual's true attribute (knowledge or ability that is being measured), but this value is impossible to identify because errors in measurement, represented by the error score, are always present (in the form of systematic and random error) whenever the obtained score is determined (Sax, 1997b). Crocker and Algina (1986) and Sax (1997b) asserted that the reduction of systematic and random error are critically important in assuring that the obtained score closely represents the true score for any given test. These authors stated that the mean error scores for the population of

students who take a test is always zero; therefore, the correlation between true scores and error scores is always zero as well. These assumptions are the foundation for the nature of the relationship between true score, observed scores, and error scores as proposed in classical test theory: the covariance between true scores and error scores is assumed to be zero, thus the observed score variance is the sum of the true score and error score variances. Finally, these authors assert that correlations between error scores from two test administrations taken by the same student would also be assumed to be zero in classical test theory. This assumption accounts for the "randomness" factor of error measurement: the two test administrations represent random sampling from the distribution of error scores, so there can be no correlation among them. Random error is present in all test administrations, but the effect that is has on the observed scores is essentially the same each time the test is taken, so it can be controlled statistically when students' scores are used to analyze the effectiveness of any test as a measure of the domain of interest (Crocker & Algina, 1986).

Sax (1997a) described a multi-step process useful in test construction that begins with establishing the reason for the test, followed by determining the objectives for the test; determining the best type of test items that would meet these objectives; and developing a test blueprint, defined as a table of specifications. HESI employs this multi-step process in creating each version of the E^2. Inclusion of a description of the processes for creating test objectives and the subsequent development of the test blueprint was beyond the scope of this research study; however, the design for each of these processes is broadly described by national accrediting bodies, such as the CCNE and NLNAC, and

9

the NCSBN, creators of the NCLEX test blueprints and publishers of periodic job analyses that define the behaviors and attributes of the entry-level registered nurse and practical nurse and state boards of nursing (American Association of Colleges of Nursing, 1999; National Council of State Boards of Nursing, 2002; National League for Nursing Accrediting Commission, 1999a). Ultimately, the depth of test item writers' clinical nursing knowledge most affects the quality of the test items developed for any test within the nursing content domain (Morrison & Free, 2001).

Morrison and Free (2001) described a model for test item writing and test construction that follows the general process as identified by Sax (1997a), but includes more rigorous specifications for the construction of critical thinking test items within the domain of nursing. Application of this method results in the production of test items that are usable for evaluating students' critical thinking abilities within the domain of nursing. Application of this four-criteria method by item writers and test designers has been shown to produce highly discriminating test items that reflect the content domain of interest, and incorporation of these types of items on a nursing test, such as the E^2, results in the development of a highly reliable test (Morrison & Free, 2001).

Using the Four-Criteria Model for Developing Critical-Thinking Test Items

Morrison and Free (2001) and Morrison, Smith, and Britt (1996) described a model for critical-thinking, multiple-choice test item writing that is used to construct each version of the E^2. This model consists of four criteria that test writers must follow when creating items for use in all HESI exams: (1) include rationale for each test item; (2) write questions at the application or above cognitive level; (3) require multilogical

10

thinking to answer questions; and (4) require a high level of discrimination to choose from among plausible alternatives. The first of these criteria refers to the use of a test not only as an evaluation tool, but also as a learning tool for nursing students.

A well-written rationale statement prepared by the test item writer and presented at the conclusion of the test administration assists students in analyzing how the correct answers were determined. This review process for all test items can be very helpful in teaching students how to examine their own thinking processes, particularly if they were unable to choose correct answers to the items during the test administration. This criterion reflects the application of critical thinking theory to nursing test item writing because individual examination of one's thinking process is a defining activity within the critical thinking process. E^2 test items are grounded in critical thinking theory.

Each version of the E^2 was developed from test banks containing questions written specifically for HESI by nurse educators and clinicians from across the United States. HESI test writers used the model described by Morrison, Smith, and Britt (1996) and Morrison and Free (2001) to develop critical thinking test items that incorporate a higher-level cognitive taxonomy. The E^2 also follows the test blueprints for the NCLEX-RN and NCLEX-PN developed by the NCSBN (National Council of State Boards of Nursing, 1999, 2001a). HESI has adopted the framework of incorporating a higher-level taxonomy as described by Bloom (1956) in developing the E^2 test items banks for the purpose of creating a product similar to NCLEX, which emphasizes the use of clinical decision-making and nursing judgment by the candidate to correctly respond to test items rather than reliance on knowledge-based facts alone. The E^2 simulates NCLEX in that

11

correct responses to test items require the application of clinical decision-making and nursing judgment rather than recitation of knowledge-based facts.

There is a considerable body of literature that examines the assumption that nurses must be strong clinical thinkers. National accrediting bodies have adopted critical thinking as a component of the standards for entry-level registered nurse competency (American Association of Colleges of Nursing, 1999; National League for Nursing Accrediting Commission, 1999a). The registered nurse licensure exam, designed to validate that candidates who pass the exam hold the minimum knowledge, skills, and judgment needed to practice professional nursing at the entry level as safe, effective practitioners, contains test items which evaluate critical thinking abilities. Thus, critical thinking must be examined in the study of the predictive accuracy of the E^2.

Richard Paul, (Paul, 1987; Paul, 1990, 1994, 1999a, 1999b; Paul & Elder, 1999), the leading authority on critical thinking, has established the relevance of the topic primarily in the domain of general education, but has also addressed the concept as it applies to the nursing discipline. Nursing researchers who have looked closely at Paul's beliefs about critical thinking, and translated those beliefs into a nursing context, have shown that critical thinking skills are important attributes of the professional nurse (Birx, 1993; Brigham, 1993; Cascio, Campbell, Sandor, Rains, & Clark, 1995; Dexter et al., 1997; Facione, 1995; Facione, Facione, & Sanchez, 1994; Kataoka-Yahiro & Saylor, 1994; Miller & Malcolm, 1990).

The second criterion of the test-writing model refers to the focus of test items on applying nursing concepts. This criterion reflects the application of Bloom's (1956)

taxonomy, a system that assists educators in defining cognitive domains of learning (Sax, 1997a). Morrison et al. (1996) stated that critical thinking test items cannot be effective if they are written to evaluate learning only at the lower levels of the cognitive domain that are characterized by rote memorization. Instead, these items should be written at a higher cognitive level, using verbs that reflect knowledge at the application level of the taxonomy and above.

The third criterion of the test-writing model refers to the use of multilogical thinking, a type of thinking described by Morrison and Free (2001) and Morrison et al. (1996) that is based upon critical thinking theory, to answer test items. Students must know more than one concept or fact to successfully answer critical thinking test items. If only one fact or concept was needed to determine the correct answer, the student would be employing memorization only, which implies development of this test item at the lower level of Bloom's taxonomy. However, multiple concepts must be linked together to create a multilogical thinking test item, and Morrison et al. (1996) suggested that the preferred method for challenging students to think critically in answering these test items involves the application of several nursing concepts in a clinically-oriented situation.

The fourth criterion of the test-writing model involves incorporation of a high level of discriminating judgment to answer the test question. Morrison et al. (1996) suggested that test item writers use only plausible alternatives to construct item distractors, implying that while all answer choices could conceivably be possible, only one is actually better than the others, i.e. with best defined as "most important, first, highest priority, and so forth…" (p. 20). The use of highly discriminating, plausible

13

alternatives when constructing multiple-choice tests promotes critical thinking among students.

Following construction of test items, Sax (1997b) suggested that educators complete four remaining steps in the test construction process: determining the scoring criteria; sequencing test item presentation; defining administration procedures, and finally, administering the test. The processes used for test construction conclude at this point, and evaluation processes begin with the first administration of the test. Sax (1997b) described criteria useful for evaluating items, tests, and measurements using concepts from classical test theory. Because error exists in all measurements produced by a test, the test use must determine how much error can be tolerated for a given use of a test, and the test designer must use reliability and validity indices to communicate the inaccuracies of the test instrument provided for the user. Reliability and validity estimates are particularly helpful for users of standardized examinations because they allow the test user to specify a degree of error that is acceptable, and choose a test that best meets the user's needs without exceeding the degree of error specified (Sax, 1997b).

To determine the effectiveness of a test in generating obtained scores that closely match true scores of subjects with repeated uses of a test within the framework of classical test theory, an estimate of reliability, or the ratio between the true variance and obtained variance, must be computed (Sax, 1997b). Because true variance is impossible to compute directly, statistical correlations that estimate this ratio are determined through the calculation of the square root of the reliability coefficient. The basic assumption of reliability coefficient interpretation is that higher correlations which approach the perfect

14

positive ratio of +1.00 reflect a closer relationship between true and obtained scores, while lower correlations that approach 0 indicate little relationship between these scores. Tests are termed "reliable" when their measurements are consistent, i.e. they dependably reflect a close relationship between true and observed scores at every administration. High error variance signals a decrease in reliability; thus, test developers strive to attain high reliability of measurements made with their tests by reducing sources of random and systematic error reflected in students' scores. Reliability coefficients are calculated through measures of stability (correlating test scores of the same students over time); equivalence (correlating scores on two or more forms of a test administered to the same students at approximately the same time): stability and equivalence (administration of two or more parallel forms of a test with a long delay between administrations); and internal consistency (correlating items on a single test (Sax, 1997b). The Kuder-Richardson Formula 20 (KR-20) is a method used to establish internal consistency from a single administration of a test (Crocker & Algina, 1986; Sax, 1997a). HESI test designers use the KR-20, as well as item difficulty, item discrimination, and other item analysis measures defined within the HPM, to: (1) compute HESI prediction scores, (2) report reliability findings on current versions of the E^2 to nursing faculties purchasing the exam, and (3) refine and improve the E^2 each time a new version is created.

The KR-20 assumes that items on the test measure the same attributes, and correlates dichotomously–scored items on the test with each other (Crocker & Algina, 1986; Sax, 1997a). When the items on the test highly correlate, the true score is reflected, meaning that error has been reduced within the observed score. The numerator of the KR-

20, a measure of the covariance among items, represents the true variance, and is divided by the obtained variance, to create the reliability coefficient (Crocker & Algina, 1986; Sax, 1997a). High KR-20 values are assumed to reflect high correlation among items, reflecting a high degree of internal consistency for the test administration, and thus, true variance (Sax, 1997b). Test developers strive to obtain high KR-20 values, which reflect high reliability coefficients, on each administration of their tests. However, identification of the best methodology to ensure the creation of highly reliable tests varies among psychometricians. In fact, test designers are encouraged to conduct item analyses using discrimination indices, such as the point biserial coefficient and item difficulty indices, such as the item difficulty level, in conjunction with reliability coefficient calculations to improve the overall quality of their tests (Sax, 1997b).

Item analysis and reliability assessment are useful in evaluating the effectiveness of any test, but they are not the only measurements of sound test construction as described in classical test theory. Crocker and Algina (1986) stated that test construction requires not only the identification of the types of behaviors or psychological constructs to be measured by the test, but also the validation of these concepts before the test can be considered useful in measuring these concepts. Concept validation is one of the keys to successful item-writing and test construction as specified within classical test theory. Validity is determined by an assessment of content validity, construct validity, and criterion-related validity. Content validity refers to the test items' effectiveness in measuring students' basic nursing knowledge and skills. Expert nurse educators and clinicians established content validity for the E^2 by evaluating the test items' relevance to

16

entry-level practice. Construct validity refers to the extent to which a test measures

specified traits or attributes at an abstract level. As a comprehensive exit exam, the E^2

measures constructs that are essential to entry level nursing practice as defined by the

NCSBN job analysis studies (Kane, Kingsbury, Colton, & Estes, 1986) and reflected in

the NCLEX test plans (National Council of State Boards of Nursing, 1999, 2001a).

Criterion-related validity refers to inferences made from analyses of students' E^2 scores

for the purpose of predicting NCLEX success. Annual research studies that correlate E^2

scores with actual NCLEX outcomes offer further evidence of the exam's predictive

validity.

It has been argued that construct validity principles as described in classical test

theory prior to the 1990's were too narrowly interpreted (Sax, 1997a). A broader view of

validity that incorporates content-related, criterion-related, and construct validity into a

single entity known as construct validity has been proposed to be more useful in

evaluating modern tests. Messick (1989, 1994) and Kane (1992, 1994, 1997; Kane,

Crooks, & Cohen, 1997) have asserted that construct validity encompasses the totality of

validity required of any test. This unitary view of validity, or unified position, supports

the assertion that construct validity exists at the very center of all strategies designed to

validate a test (Sax, 1997b). The emphasis on the meaning of the test scores and how they

are interpreted has replaced previously held beliefs about the establishment of validity,

and is central to the unified position of validity (Sax, 1997a). While Kane (1994) stated

that even the unified view might not be adequate to explain the validity of professional

licensure examinations, he concurred with Messick (1989) on the premise that test validity rests predominantly upon the evidence of construct validity.

Application of the unified view of validity has been used by Kane (1994) to demonstrate the predictive validity of licensure examinations on future professional practice through the process of constructing arguments using empirical evidence as a basis to defend, or validate, the usefulness of a test for a specified purpose. Construction of an argument asserting that the E^2 is predictive of success on NCLEX using the unified view of validity as conceptualized by Kane (1994) is useful in establishing construct validity for the E^2. It has been asserted that empirical data demonstrating the link between test score and professional practice criteria are needed to validate licensure examinations (Kane, 1992). Thus, this view of construct validity would indicate that an individual's performance on a licensure examination simulation test can be predictive of: (1) performance on the licensure examination of interest and (2) the professional behavior(s) of interest (Kane et al., 1986).

Michael Kane has been a frequent psychometric consultant of licensure examination publishers for the health professions and has extensively examined the use of licensure examination results as predictors of candidate performance in professional practice (Kane, 1994; Kane et al., 1986). The NCSBN, developer of the NCLEX licensure exams, employs Kane's view of content validity in creating NCLEX test blueprints, which makes Kane's process of identifying arguments that support the use of a particular test for a particular purpose useful in establishing the predictive validity of the E^2 on NCLEX outcomes (Kane et al., 1986; Smith & Crawford, 2002).

18

The NCSBN conducts detailed practice analyses, the RN and PN Job Analyses, every three years for the purpose of supplying the NCSBN with knowledge about the practices of newly-licensed nurses so that these findings can be directly linked to NCLEX-RN and NCLEX-PN content (National Council of State Boards of Nursing, 2002; Smith & Crawford, 2002). The latest NCSBN RN Job Analysis was conducted in 1999, and the data were used to validate that the NCLEX-RN exam blueprint included proportionately more content on job activities that were rated as high-priority, high frequency activities by entry-level registered nurses (National Council of State Boards of Nursing, 2002). A test, such as the E^2, that simulates the NCLEX-RN and follows the NCLEX-RN test blueprint would necessarily reflect the same proportions of high-priority, high frequency nursing activities as a demonstration of validity. Kane argues that because the RN and PN job analyses conducted by the NCSBN are used to construct the test blueprint for the NCLEX exams, these licensure exams in turn reflect the knowledge, skills, and abilities used by entry-level nurses functioning within these jobs, and are therefore valid tests for predicting which candidates possess the nursing competencies required of those who are to assume entry-level employment as RNs and PNs, and which do not possess the required nursing competencies.

Assumptions

Assumptions associated with classical test theory relate to beliefs about the nature of psychological attributes and human behavior and capabilities for measuring these in a meaningful way. Classical test theory asserts that psychological attributes can be measured by assigning quantitative values to behaviors that are reflective of the attribute

19

of interest. Constructs can be used to identify behaviors and attributes, and these are also measurable. Measuring behaviors and attributes defined by a group of constructs can also be used to predict behavior in a future situation. When behaviors are measured on a test, the observed score in combination with the error score reveals an individual's true score, which is assumed to best describe that individual's attribute of interest. Controlling error through deliberative test construction increases the probability that the true score and the observed scores will correlate.

Statistical methods can be applied to evaluate the effect of error on the student's true score on a given test instrument, and result in measures of the test's reliability. Further, the content validity, construct validity, and criterion-related validity of a test can be determined. A unified view of content validity that encompasses the three types of validity as explained by classical test theory can also be constructed through inference or argument, and statistical evidence can be accrued to support the inferences made or arguments constructed to assert the validity of a test for a given purpose. The use of argument as a means of determining construct validity for tests designed to possess a high degree of predictive accuracy regarding performance on a professional licensure examination, and ultimately, regarding competence in an entry-level job, is an accepted practice in the profession of nursing. Specific assumptions associated with a unified view of content validation for licensure exams and tests that simulate licensure exams center on the construction of the arguments or inferences about the usefulness of a test as a measure of the criterion of interest. Tests designed to simulate a licensure exam constructed to predict competency in entry-level nursing jobs can be validated using a

carefully constructed argument asserting that a test is reflective of the high frequency, high-priority behaviors inherent in the entry-level job as determined through job analysis.

In summary, the main assumption of this study was that a deliberatively constructed comprehensive nursing exam, with established validity and reliability, can yield predictive information regarding critical thinking ability and predicted performance on a national licensure exam (Facione, 1995, 1999a, 1999b; Facione & Facione, 1996a, 1996b; Facione et al., 1994; Paul, 1994).

Research Questions

The following research questions were addressed in this study:

1. What is the predictive accuracy of the E^2 for students in associate degree, baccalaureate degree, and diploma RN programs and PN programs who took the test 4 to 6 months prior to graduation in academic year 1999-2000, whose prediction scores fell into one of five scoring interval categories?

2. What criteria used to determine readiness for graduation are articulated in the policies of schools administering the E^2, including the identified E^2 benchmark for progression in a nursing program?

3. What remediation strategies are employed by schools of nursing administering the E^2 to assist students who fail to reach established minimal E^2 benchmarks?

Definition of Terms

The terms used in this study were defined as follows:

1. Students were defined as individuals who took the E^2 for the first time during academic year 1999-2000, identified as Year IV, which was the fourth consecutive

21

year of study conducted on the E^2, within four to six months prior to graduation from an RN or PN school of nursing.

2. Prediction scores were the total scores on the E^2 that were calculated using the HESI Predictability Model (HPM), a proprietary mathematical model used to describe the student's probability of passing the licensure examination. All HESI probability scores are calculated using the HPM. HESI scores are not expressed as percentages. Instead, these scores are calculated by applying the HPM to the raw score. The HPM considers the difficulty level of each test item in determining students' performance on the exam.

3. HESI scoring categories were used on the HESI scoring reports, and these ranged from A, the highest-scoring category, to H, the lowest-scoring category. These scoring categories served as the basis for formulating the HESI scoring intervals used for data collection and data analysis. Previous studies examined only high-scoring students, defined as those who scored in categories A and B (90-99.99), and low-scoring students, defined as those who scored in categories G and H ($\leq$69.99). In this study, additional scoring intervals were designated to provide more discrimination in the middle-scoring groups where the greatest ambiguity existed regarding the degree of risk for NCLEX failure.

4. Predictive accuracy was described as the percentage of occurrences of the E^2 scoring interval successfully predicting a student's outcome on the NCLEX on the first attempt.

5. E^2 benchmark scores were defined as the specific, numerical HESI prediction scores selected and reported by schools of nursing as the minimally-acceptable scores students were required to attain before progression to graduation and approval for NCLEX candidacy were granted by the nursing faculty and administration. These scores were included in schools' progression policy statements.

6. Remediation strategies were the reported programs of instruction including elective courses, computer assisted instructional packages, NCLEX review books, and other educational materials used by students who failed to attain required E^2 benchmark scores. These programs and materials were used by students and faculty to assist students in improving up weak content areas identified in the E^2 scoring reports.

Limitations

Certain limitations of the research methodology and analytic techniques must be considered when reviewing the findings of this study.

1. Incorporation of more sophisticated computerized tracking techniques into the data acquisition process rendered some subjects' scores inaccessible for this study. Therefore, despite the fact that approximately 50% more E^2s were administered in Year IV (academic year 1999-2000) than in Year III (academic year 1998-1999), the Year IV sample size was only slightly larger than that of Year III.

2. The research design lacked control over intervening factors that may have occurred between E^2 administration and administration of the NCLEX-RN or NCLEX-PN. Such factors could have influenced the predictive accuracy of the E^2.

Summary

The fourth annual validity study of the HESI E^2 was designed to examine not only the accuracy of the E^2 in predicting NCLEX success, but also the degree of risk for failure of the licensure exam associated with specific scoring intervals. A theoretical framework incorporating elements of critical thinking theory and test (psychometric) theory was used to establish the reliability and validity of the E^2 as a predictor of NCLEX outcomes. In addition, the study was undertaken to discover the criteria used by nursing faculty to determine their students' readiness for graduation that incorporates E^2 results as benchmarks, and discover the types of remediation strategies employed to assist low-scoring students who fail to achieve the required E^2 benchmark. Limitations to the study design include use of a convenience sampling method that resulted in the exclusion of some of the potential subjects' scores and lack of control for intervening variables occurring between E^2 administration and NCLEX administration that may have influenced NCLEX outcomes.

CHAPTER 2

REVIEW OF LITERATURE

For more than forty years, nurse educators have searched for predictors of

NCLEX success. However, despite the plethora of nursing research studies conducted

regarding predictors of NCLEX success, dependable predictors of licensure success

remain elusive. In addition, changes in the characteristics of applicants to nursing

programs in recent years have confounded the search for previously reported predictors

of NCLEX success that held promise.

A review of current literature on historic and current predictors of NCLEX

success, the characteristics of nursing program applicants; benchmarking for progression

within nursing education programs; and the design and implementation of remediation

strategies are presented. Additionally, findings of previous validity studies conducted that

focused on the instrument under study, the HESI E^2, are described. The intent of this

review of the literature is to offer background information about the variables of interest

in this study. The review also offers support for continuing the search for the elusive

predictor variables related to licensure success and further study of the E^2 as one such

predictor.

Literature Reviews Summarizing Forty Years of Prediction Studies

So much has been written on the subject of predictors of licensure success that several

authors have published literature reviews summarizing the findings. Taylor et al. (1966;

1965) reviewed the literature on predictors of licensure success conducted prior to 1965. Of 180 studies found, 77 were considered to be prediction studies in basic and graduate nursing programs. Five major criteria and approximately 65 predictors of success of students in nursing programs were identified. The authors identified the major criteria, their frequency of use within the 77 studies, and the ranges of correlations with other predictors: (1) Grades, 51, .-24 - .65; (2) State Board Examination, 36, .00 - .63; (3) Survival in School, 17, ..03 - .47 (this criterion reflects student attrition); (4) Ratings, 13, -.25 - .40 (this criterion refers to subjective assessment of students, primarily through clinical evaluation measures devised by the instructor); and (5) Job Performance, 2, -.15 - -.19 in the first study, and -.02 - .21 in the second study. The authors criticized the 77 studies reviewed, stating that less attention overall was focused on predictors of success of nursing students in their performance as registered nurses than on their performance as students. They also expressed concern regarding the lack of nursing research in relating selection criteria for nursing school applicants to success in nursing practice, and suggested that tests in use for nursing applicant selection be validated against job performance criteria. Taylor et al. (1966) found that most predictors used by schools of nursing related to standard indicators of scholastic aptitude or a student's record of academic achievements while in nursing school.

In 1978, Schwirian, Baer, Basta, and Larabee (1978) published an updated literature review on predictors of licensure success that examined data published from 1965 to 1975. These authors cited 12 studies that examined predictors of students' success in nursing school. Seven studies focused on predictors of students' success on the State

Board Test Pool Examinations (SBTPE), the licensure examination that was a predecessor to the NCLEX. Eight prediction studies on job success after completion of nursing school. The authors reported that the primary predictors of student success on the SBTPE were NLN Achievement Tests scores, theory course grades, and GPAs, and that grades in students' clinical courses were not predictive of licensure exam performance. The authors did not indicate statistically how they arrived as the conclusions about the best predictors of success, but instead summarized the sample size, type of predictor, measure used for the predictor, and descriptions of the findings of each study in a table within their report. Similar descriptive information was supplied about each of the seven studies that identified predictors of successful job performance after graduation. The authors concluded that nursing clinical grades were the best predictors of performance in nursing practice, which represented a complete reversal from the predictive value assessed for clinical grades with regard to the SBTPE.

Carpenter and Bailey (1999) updated the findings of Taylor et al. (1966) and Taylor et al. (1965) and Schwirian et al. (1978) when they examined the literature regarding predictors of licensure success published between 1976 and 1998. A total of 67 studies were included in the review by these authors. In studies involving baccalaureate nursing students, several authors employed stepwise logistic regression as the multivariate technique to establish the usefulness of specified prediction variables. The authors concluded that logistic requesting was an appropriate technique for this type of analysis because it does not require assumption of normal distributions for NCLEX-RN pass rates or GPA or ACT scores. In addition, they stated that descriptive statistics,

Pearson correlation coefficients, and varimax rotated factor analysis were also used to analyze predictive value of GPAs, SAT and ACT scores, and outcomes on standardized tests such as NLN examinations and the Mosby Assess Test. With regard to the analysis of associate degree students' predictors of success on the NCLEX-RN, the authors found that multiple regression and correlation analysis were also used. Their overall findings indicated that: (1) associate degree (ADN), baccalaureate (BSN), and diploma programs shared similar predictors of NCLEX success; (2) academic factors and high school rank positively correlated with NCLEX success; and (3) nursing theory courses in combination with NLN test scores appeared to be the best predictors of success. They also reported that between 1976 and 1998, the NLN Baccalaureate Achievement Test and the Mosby Assess Test were the exams most frequently used to predict NCLEX success. Table 1 summarizes selected studies in alphabetical order by author identified by Carpenter and Bailey (1999) and findings of studies released since this review was published that analyzed the accuracy of selected predictor variables. The type of predictor variables chosen, the statistical tests used for analysis, any intervention given to subjects, and the findings relevant to predictive accuracy on NCLEX outcomes are shown in this table.

Despite the plethora of nursing research studies conducted between 1976 and 1998 regarding predictors of NCLEX success, these authors concluded that a consistently stable or reliable predictor of licensure success had not yet been described.

Table 1.

Summary of Studies Analyzing Predictor Variables of NCLEX Success

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
Alexander & Brophy (1997)	High school rank, SAT scores, Chemistry & Math, taken in high school, admission status, GPA, age, theory courses, NLN Comprehensive Achievement Test	188 ADN graduates	t-tests, chi-square, logistic regression	None	SAT verbal scores, nursing GPA and NLN Comprehensive Achievement Test Scores ($t(128)=8.08$, $p<0.0001$).

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
Arathuzik & Aber (1998)	Study Skills Self-Efficacy Instrument (SSSE)	79 BSN students	Mean scores, standard deviations, and ranges of scores; point biserial correlations between variables identified on the SSSE and NCLEX-RN outcome	None	Significant correlations (p=.05) for passing the NCLEX-RN were found with: nursing GPA (PBCC .275), English as the first language (PBCC .253) , lack of family responsibilities (PBCC -.293), lack of emotional distress (PBCC -.240), and perception of competency in critical thinking skills (PBCC .245).
Ashley & O'Neill (1994)	Cumulative GPA, nursing GPA, SAT verbal and math	125 BSN students	t-tests, chi-square	Participation in NCLEX study groups	NCLEX-RN pass rate for study group participants was 8% higher than non-participants, but the increase was not

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	scores; Mosby Assesstest scores				statistically significant.
Ashley & O'Neill (1991)	Cumulative GPA, nursing GPA, SAT verbal and math scores; Mosby Assesstest scores	30 BSN students	t-tests, chi-square	Participation in test-coaching course	NCLEX-RN pass rate for students at-risk of NCLEX-RN failure (χ^2=6.30, p=.01) and not-at-risk (χ^2=13.87, p=.001) that participated in a test-coaching group achieved significantly better NCLEX-RN pass rates than non-participants.
Barkley et al. (1998)	NCLEX-RN Risk Appraisal Instrument (NCLEX-RN RAI); nursing course grades; NLN Adult	81 BSN students	Mean scores, standard deviations, and ranges of scores; Mann-Whitney U; chi-square;	None	Significant correlations (p=.001) were found between NCLEX-RN performance and performance in a pediatric nursing course (r=.5873); a psychiatric nursing course (r=.5825) and the NLN Adult Achievement Test (r=.5851). Findings of these

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	Achievement Test		Pearson correlations		correlations were applied to the NCLEX-RN RAI, and it was determined that a score of 0 indicated virtually no risk of failing the NCLEX-RN, but scores of 1 to 11 or greater resulted in minimal to certain risk of failure.
Beemon & Waterhouse (2001)	Year of graduation, sex, age at graduation, type of program, SAT verbal and math scores, biology, physiology, and pathophysiology grades, nursing	289 BSN students	Pearson correlations; discriminant analysis	None	Total number of C+ or lower grades in nursing theory courses ($r=-.394$, $p<.0001$) and grades in two additional nursing courses ($r=.381$, $p<.0001$; ($r=.379$, $p<.0001$).

32

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	course grades, mid-curricular GPA and senior year GPA				
Beeson & Kissling (2001)	Type of student, age at time of the licensure exam, gender, performance in pre-nursing courses, cumulative GPA at graduation, and Mosby Assesstest score	505 BSN students	*t*-test, Mann-Whitney, chi-squares, logistic regression	None	Number of Cs, Ds, and Fs in junior year nursing courses were significantly related to NCLEX-RN results (χ^2=108.42, *p*=.0001); logistic regression model correctly predicted 76% of the students who failed based on information available on students by the end of the first semester of the senior year.
Billings et al.	Eight computerized	None	Frequencies	None	Critique offered on each of the

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
(1996)	NCLEX-RN preparation programs, including the HESI Test, were compared		associated with evaluation criteria, i.e. program design, question types and feedback, performance reports; screen design		selected computerized programs.
Brown (1987)	Grade in first year nursing course	26 BSN students	*t*-test	Support group	Significant difference ($p=.05$) between NCLEX scores of at-risk students who did not participate in a support group compared to at-risk

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
					students who participated in a support group ($t=2.10$).
Campbell & Dickson (1996)	Cognitive predictors (cumulative college GPAs; GPAs of courses, including those in prenursing, nursing, biology, chemistry, liberal arts, social psychology, and mathematics; SAT, ACT, and NLN exams; Watson-	Integrative review of 47 studies and meta-analysis of 4 studies over a ten year period, 1981-1990	Descriptive quantification of study characteristics for integrative review and chi-square, z, weighted average d, binomial effect size display (BESD),	Support group, computer-assisted instruction, personalized system of instruction, and integrated curriculum	All interventions had some significance in between-group variations, but only the support group was predictive ($p=.05$) of NCLEX success ($t=2.10$).

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	Glaser Critical Thinking Appraisal; Mosby Assesstest. Self-enhancement variables were cognitive/learning style, self-concept/esteem, test anxiety, social support,situational.		homogeneity analysis (Q), and fail-safe N (N_{FS})		
Endres (1997)	Admission GPA, medical-surgical nursing GPA, nursing GPA, Mosby Assesstest,	50 African American, 50 foreign-	Correlations, chi-square, two-way ANOVA, discriminant	None	No significant differences were found between NCLEX-RN pass rates among African American, foreign born, and white graduates. Students in all three sampling groups who had

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	age, current placement within the curriculum, PN licensure, and number of Ds and Fs received in nursing courses	born, and 50 white BSN students	analysis		Mosby Assesstest percentile rankings below 21 $\chi^2(1,50)=9.09$, $p=.004$ and a D or F in a nursing course ($p=.001$) were more likely to fail the NCLEX-RN than those with higher percentile rank and no D/F.
Frierson (1993)	None	8 BSN students at historically black state-supported university	Correlation	Three-pronged intervention procedures consisted of test-taking skills, group learning, and instructor reinforcement	NCLEX-RN pass rate for students that received the three pronged intervention was significantly ($z=1.97$, $p=.025$) higher than for students who did not receive the intervention.
Heupel (1994)	GPAs at the end of	152	Pearson	None	Grades in nursing theory courses

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	each of the four collegiate years; nursing theory course grades	baccalaureate students	correlations, multiple regression analysis		$(r=.604, p=.0001; (r=..493, p=.001; (r=.716, p=.01)$; junior year GPA $(r=.674, p=.001)$.
Lauchner et al. (1999)	HESI Exit Exam	2809 students in RN and PN programs	Chi-square	Monitoring of exam administration	No significant difference found in the predictive accuracy of the HESI Exit Exam for students in all types of nursing programs (ADN, BSN, Diploma, and PN). The exam was significantly (χ^2=4.98, p=.05; more accurate (99.49%) when administration was monitored versus unmonitored (96.82%).
Lockie & Burke (1999)	Cumulative GPA, Mosby Assesstest	210 at-risk BSN students	Chi-square	Enrollment in remediation course entitled Partnership	Significant differences found in retention (29.9%, p=.00001) , graduation rates (50.4%, p<.00001),

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
				in Learning for Utmost Success (PLUS)	and attrition rates (9.9%, $p<.00001$) between at-risk PLUS students and at-risk non-PLUS students. GPAs, Mosby Assesstest scores, and NCLEX-RN pass rates were not significantly different between these groups.
Marshall (1989)	Malone Social Network Inventory (MSNI)	161 ADN students	Mean scores, ranges of scores	None	Students who persisted in a nursing program had significantly more "great impact" relationships among their network members according to the MSNI than students who dropped out of the program. Age was the most distinguishing demographic variable in that students who persisted in the

39

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
					program had a mean age of 27.8 years, while greater attrition was noted among students with a mean age of 33.5.
Memmer & Worth (1991)	30 remedial strategies analyzed: placement tests, remedial courses, medical/nursing terminology courses, career day seminars, preadmission meetings with instructors,	21 BSN program directors	Frequencies, comparisons, Sign Test	30 retention approaches aimed at reducing attrition rates among English as a second language students (ESL),	The five programs with highest ESL retention rates were compared with the 21 total programs on 47 retention strategies using the Sign Test. In 40 out of 47 comparisons, programs with the highest ESL student retention rates reported using the greater percentage of the 30 retention approaches explained in this study ($p<.01$).

40

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	orientation programs, study skills workshops, writing workshops, outreach to students' families, employment of nurse-coordinator to monitor ESL students' progress, mentor programs, employment of culturally diverse faculty, flexible class load and class time options,				

41

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	reduced instructor/ student ratios in clinical, heterogeneous ethnic grouping of students in clinical, additional clinical and skills lab assistance, peer tutorial assistance, academic advising, housing with other ESL students, ESL student participation in campus				

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	organizations, ESL faculty invited as guest speakers, improved communication with ESL students, sensitivity training for faculty, early intervention				
Newman et al. (2000)	HESI Exit Exam	3,752 students in RN and PN programs	Chi square	None	No significant difference found in the predictive accuracy of the HESI Exit Exam for students in all types of nursing programs (ADN, BSN, Diploma, and PN). Significantly more (χ^2=488.08, p=.01) RN students who had low-scores on the HESI Exit

43

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
					Exam failed the NCLEX-RN. Significantly fewer (χ^2=6.46, p=.05) RN students who were from schools that used the HESI Exit Exam for remediation failed the NCLEX-RN than those from schools that did not use the exam for remediation. The exam was significantly (χ^2=.017, p=.001) more accurate (98.74%) when administration was monitored versus unmonitored (97.20%).
Nibert & Young (2001)	HESI Exit Exam	6,277 students in RN and PN programs	Chi square	None	No significant difference found in the predictive accuracy of the HESI Exit Exam for students in all types of nursing programs (ADN, BSN, Diploma, and PN). Significantly more

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
					(χ^2=818.775, p=.01) RN students who had low-scores on the HESI Exit Exam failed the NCLEX-RN. Significantly fewer (χ^2=6.46, p=.05) RN students who were from schools that used the HESI Exit Exam for remediation failed the NCLEX-RN than those from schools that did not use the exam for remediation. The effect of monitoring and the use of remediation on students' NCLEX-RN pass rates were not significant.
Poorman & Martin (1991)	Test anxiety inventory, cognitive assessment tool,	102 BSN students	Multiple regression, Pearson correlation,	None	Test anxiety was inversely related to NCLEX-RN success. Academic aptitude was positively correlated to NCLEX-RN success. Negative

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	quality point average, SAT scores		chi square		cognition was not inversely related to NCLEX-RN pass rates. Predictions of grades and predictions of NCLEX-RN outcomes by the student were the best predictors of actual NCLEX-RN scores.
Riner et al. (1997)	Nine computerized NCLEX-RN preparation programs, including the HESI Test, were compared	None	Frequencies associated with evaluation criteria such as program design features, types of questions and	None	Critique offered on each computerized program.

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
			feedback, performance reports, and screen design		
Ross et al. (1996)	Computer-Assisted Preparation for the NCLEX-RN	230 BSN students	Mean scores and ranges on practice exams compared with NCLEX outcomes	None	Mean scores on practice exam for students who failed the NCLEX-RN were 57.25% (rage, 50-61%). Mean scores on practice exam for students who passed the NCLEX-RN were 66.81% (range, 59%-80%).
Schmidt (2000)	NLN Diagnostic Readiness Test and Pre-Admissions test scores	5,698 students in RN programs	Mean, standard deviation, and ranges of DRT and	None	Centered DRT scores (scores centered about the grand mean) were significant ($B=.0476$, $p<.001$) predictors of NCLEX outcomes. While support was found for the

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
			Pre-Admission scores; chi-square, hierarchical logistic regression model		predictive validity of the DRT, the same was not found for the Pre-Admission test scores.
Wall et al. (1993)	SAT scores, high school rank, NLN Achievement test scores, Mosby Assess Test scores, theory GPA	92 BSN students	t test, discriminant function analysis	None	The nursing GPA was the most accurate predictor of NCLEX outcomes (F=21.047(1, 92), p=.001.)
Waterhouse et al. (1993)	SAT verbal and math scores; high	257 BSN graduates	Means, standard	None	Graduation GPIs showed the highest correlation with NCLEX success

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	school percentile, physiology grades, nursing course GPAs, sophomore and graduation grade point indexes (GPIs)		deviations , frequency distributions, Pearson correlations		(r=.248, $p\leq$.05). Nursing GPA (r=.216, $p\leq$.05) and SAT verbal scores (r=.232, $p\leq$.05) were also highly correlated with NCLEX-RN success. The discriminant analysis correctly categorized 91.44% of the subjects regarding NCLEX outcome.
Waterhouse (1994)	SAT verbal and math scores; high school percentile, physiology grades, nursing course GPAs, sophomore and graduation grade point indexes (GPIs),	142 BSN students	Means, standard deviations, and frequency distributions, Pearson correlations, t-test, chi		Cross-validation with data from the 1993 study revealed that scores were significantly lower for the current study subjects than for study subjects in the previous year: SAT verbal (t=2.675, $p\leq$.01); SAT math (t=3.004, $P\leq$.01); high school percentile (t=3.069, $p\leq$.01); physiology grades (t=3.243, $p\leq$.01); nursing theory

Author/Date	Predictor Variables	Sample	Statistical Methods	Interventions	Findings: Best Predictors
	participation in American Nursing Review (ANR) course		square, discriminant function analysis		grades (t=7.780 $P\leq$.01; t=2.675, $p\leq$.01). There were no significant differences between the two samples in numbers of students passing the NCLEX-RN. The discriminant analysis correctly categorized 84.00% of the subjects regarding NCLEX outcome.

Predictive Accuracy

Although consistently stable predictors have failed to appear within the past 40 years, nursing educators have nevertheless experienced limited successes with a few well-known predictors used to identify students at risk of failing NCLEX. Campbell and Dickson (1996) completed an integrative review and meta-analysis that included the findings of 47 studies examining predictor variables for NCLEX conducted between 1981 and 1990. Results of the review by these authors showed that science course grades and GPAs in nursing courses were the greatest cognitive predictors of NCLEX success and that parental education levels and the age of students comprised the greatest demographic predictors of success. Ninety-four percent of the studies reviewed were descriptive, used convenience samples, and most used quantitative measures as predictors. The meta-analysis demonstrated significant effectiveness of several educational and psychological interventions undertaken in four studies. While the studies included in the study by Campbell & Dickson (1996) focused on graduates taking the NCLEX prior to the implementation of the 1988 exam revision, which raised the passing standard and reported outcomes on a pass/fail basis , and prior to the 1994 adoption of the computerized adaptive testing (CAT) format, these same predictor variables have continued to be studied.. In fact, several authors confirmed the association between academic factors, such as admission criteria, including the Scholastic Aptitude Test (SAT), and American College Test (ACT) scores, high school percentile rank and pre-nursing science and liberal arts course grades, as well as nursing grades, with NCLEX success (Alexander & Brophy, 1997; Barkley, Rhodes, & Dufour, 1998; Beeman &

Waterhouse, 2001; Beeson & Kissling, 2001; Endres, 1997; Frierson, Malone, & Shelton, 1993; Heupel, 1994; Vance, 1997; Wall, Miller, & Widerquist, 1993; Waterhouse, Bucher, & Beeman, 1994; Waterhouse, Carroll, & Beeman, 1993).

Both Waterhouse et al. (1993) and Wall et al. (1993) determined that predictors of NCLEX-RN success post-1988 differed from the pre-1998 examination. The study conducted by Waterhouse et al. (1993) studied 257 students in a baccalaureate nursing program who graduated between 1988 and 1990. Using discriminate analysis, investigators demonstrated that grades achieved in the first senior year nursing course and the GPA achieved by the end of the nursing program were the most predictive variables of NCLEX-RN outcomes. Waterhouse et al. (1994) continued the work begun in the 1993 study when they conducted a cross-validation of student data obtained between 1991 and 1992. These authors found that eighty-four percent of the students' NCLEX outcomes were correctly predicted using the first senior nursing course grade and end of Program GPA as Waterhouse et al. (1993).

Non-academic factors such as age, emotional state, ethnicity, family responsibilities, non-English primary language, stress, self-esteem, time management, and test anxiety have also been related to NCLEX success, but they have generally been less predictive than academic factors (Arathuzik & Aber, 1998; Billings et al., 1996; Endres, 1997; Poorman & Martin, 1991). Arathuzik and Aber (1998) focused on the cultural diversity of undergraduate nursing students in 1998 and recognized that non-academic predictors of NCLEX needed to be studied to determine their influence within a student population that included many different ethnic groups, many students for whom

52

English was a second language, and encompassed a wide range of ages. This descriptive, correlational research included 79 generic senior students in a nursing program within an urban university and revealed that several internal and external blocks to success. External blocks to success included burdens related to jobs, family, and financial responsibilities, while internal blocks were associated with emotional stress and fatigue that stemmed from multiple role strain while pursing nursing studies. These blocks negatively impacted NCLEX-RN success. Positive correlations were found between NCLEX-RN success and the nursing program GPA, critical thinking ability, and reduced demands from family responsibilities. While these variables showed low correlations ($r =$ .24 - .28), they were indeed significant ($p = .05$).

Previous studies have compared the effectiveness of these predictor variables not only for traditional students, but also for the increasing numbers of non-traditional students in nursing, including minorities and English-as-a-second language students. These students have documented difficulties in gaining entrance, maintaining persistence, and achieving success on the NCLEX-RN. Findings indicated that a lack of academic preparedness, particularly in the sciences and English; poor career preparation offered by high school counselors; financial constraints; and a lack of adequate scholarships negatively affected those students' ability to succeed in college. Additionally, minorities entering nursing schools reported feelings of social isolation, alienation, and racism, including a pattern of institutional racism espoused by many universities (Allen et al., 1988; Rosella, 1994; Tucker, 1999). However, recruitment of culturally-diverse students has been recommended for many years as one strategy to resolve the nursing shortage.

53

(Allen, Nunley, & Scott-Warner, 1988; Barbee, 2001; Grossman et al., 1998; Rosella, 1994). As minority and ESL student cohorts increased within the nation's nursing schools over the past few years, faculties began to recognize the necessity of offering these students enhanced support to reduce their attrition rates and enhance their prospects for NCLEX-RN success after graduation.

Faculties responded to concerns about rising attrition and the risk of decreasing NCLEX-RN pass rates by designing programs aimed at retaining culturally-diverse students who were known to be at greater risk for attrition and NCLEX failure than non-minority, English-speaking students (Arathuzik & Aber, 1998; Endres, 1997; Frierson et al., 1993; Tucker, 1999). To combat rising attrition rates among culturally-diverse students and reverse the trend of declining NCLEX pass rates, nursing faculties structured formal and informal support programs intended to bolster academic skills, address nursing knowledge deficits, and facilitate social support systems within the collegiate environment (Burris, 1987; Frierson et al., 1993; Lockie & Burke, 1999; Marshall, 1989; Memmer, 1991; Rodgers, 1990; Wolahan & Wieczorek, 1991). Many of these remediation programs have been implemented in the final semester of the nursing program culminating in graduation, and they have focused on student weaknesses in specific NCLEX content categories.

Factors associated with the highest predictability of NCLEX success, such as cumulative grade-point average, grades in senior level nursing courses, and outcomes on NCLEX readiness tests, occur at the end of the nursing program. Obtaining information this late in a student's curriculum leaves little time for NCLEX preparation, much less

specific remediation of identified deficit areas. Nurse educators have increasingly turned to the use of standardized tests given in the later half of the student's senior year in nursing school to identify individuals who are at risk for NCLEX failure as early as possible so that remediation can be promptly initiated. Several comprehensive exams have demonstrated a moderate to high ability to predict NCLEX success (Alexander & Brophy, 1997; Barkley et al., 1998; Breyer, 1984; Endres, 1997; Ross, Nice, May, & Billings, 1996; Schmidt, 2000). In a recent study of 5,698 students in 135 schools of nursing who were administered two tests published by the National League for Nursing (NLN), the Diagnostic Readiness Test (DRT) and Pre-Admissions Test, Schmidt (2000) found that student scores on the DRT were the only scores that were significant (B=.0476, p<.001) predictors of NCLEX performance using a hierarchical logistic regression model. Most standardized tests, such as the DRT published by the NLN are paper and pencil administered tests, and results are not available soon enough to use the data as a remediation resource. Also, paper and pencil tests do not provide students with practice using the NCLEX computerized adaptive testing (CAT) format, which can be valuable in preparing students for the keystroke mechanics associated with taking licensure exams. For this reason, nursing faculties have increasingly chosen computerized instruments to help prepare students for NCLEX (Billings et al., 1996; Riner et al., 1997; Waterhouse et al., 1994; Waterhouse et al., 1993). Software companies have responded by providing a variety of computer products, with expanded use of computerized testing now becoming common within nursing schools (Billings et al., 1996; Riner et al., 1997). However, Beeman and Waterhouse (2001) reported that validation of the E^2 by Lauchner,

55

Newman, and Britt (1999) was the only study found that attempted to identify predictors of NCLEX success post-1994, i.e. since the introduction of the NCLEX version in CAT format.

HESI produces a variety of nursing exams, all of which are computerized, allowing students to receive their scores immediately upon exam completion so that a specific remediation plan can be quickly implemented if needed. Using a nationwide database, the E^2 provides an analysis of student performance that allows schools to compare their nursing program with programs throughout the United States. The E^2 also uses the same keystrokes as the NCLEX, thereby simulating the mechanics of NCLEX administration. More than one version of the E^2 is available so that those students who have been remediated can be retested to evaluate the effectiveness of their remediation programs (Lauchner et al., 1999; Newman, Britt, & Lauchner, 2000; Nibert & Young, 2001).

E^2 scores of students in associate degree, baccalaureate degree, and diploma nursing programs leading to registered nurse licensure and in practical nursing programs leading to practical nurse licensure comprised the samples for each of the E^2 studies. Successively larger sample sizes were reported in each of the three published E^2 validity studies. The total sample size for Year I was reported by (Newman et al., 2000) as 2,809 students. However, the reporting sample for Year I was only 2725 since 84 of the total did not respond. In looking at the respondents only, in Year II, the reporting sample was 3752; and in Year III, it was 6277. Almost half of all data gathered for all three years were collected in Year III. Table 2 describes the sample by years and types of programs.

56

Table 2.

E² Aggregate Data: Sample Size by Group for Years I, II, and III

Description	Year I	Year II	Year III	Total
RN	2555	3296	5588	11,439
ADN	1976	2456	3651	8083
	(77.34%)	(74.51%)	(65.34%)	
BSN	520	796	1921	3237
	(20.35%)	(24.15%)	(34.38%)	
Diploma	59	44	16	119
	(2.31%)	(1.34%)	(0.28%)	
PN	170	456	689	1315
Total RN/PN	2725	3752	6277	12,754

Findings indicated that in Year I (1996-97) the predictive accuracy of the E^2 was 97.41% (N = 2,725) (Lauchner et al., 1999), in Year II (1997-98) 96.49% (N = 3,752) (Newman et al., 2000), and in Year III (1998-99) 97.78% (N = 6,277) (Nibert & Young, 2001). In all three studies, no significant difference was found in the predictive accuracy of the E^2 regardless of the type of program administering the exam, including ADN, BSN, diploma, and practical nursing (PN) programs.

In Year II, the first year NCLEX outcomes of low-scoring E^2 students were examined, significantly more ($p = .001$) low-scoring E^2 students failed the NCLEX than

57

did high-scoring E^2 students. Furthermore, when the E^2 was used as a guide for remediation, significantly fewer ($p = .01$) of the low-scoring E^2 students failed the licensing exam than when the E^2 was not used to guide remediation (Newman et al., 2000). In Year III, as in Year II, a significantly greater incidence of NCLEX failure was documented among low-scoring E^2 students ($p = .001$) than among high-scoring E^2 students. However, no significant difference was found in low-scoring E^2 students' pass rates related to their participation in remediation programs (Nibert & Young, 2001). A summary of the findings from the three previous E^2 validity studies is present in Table 3.

Benchmarking

Benchmarking is a process for establishing an expected level of quality that originated with surveyors who marked objects in the landscape to indicate a reference point for determining altitudes (Rudy, Lucke, Whitman, & Davidson, 2001). Benchmarking first gained recognition in business and industry as a technique for quality improvement wherein companies identified "best practices" through a process of comparing performance on specific variables. Benchmarking research studies sought to identify and promote adoption of outstanding practices from one or more institutions to others to achieve a recognized goal (Billings, Connors, & Skiba, 2001). The practice of benchmarking within the health care industry of the United States has been used to identify and judge high quality care through the interpretation of patient outcome data from leading health care institutions leading to the practice of publishing hospital "report cards" periodically to designate a particular institution's track record in meeting

58

Table 3.   Summary of Findings by Year

	Year I	Year II	Year III	Summary of Findings
Accuracy	97.41%	96.49%	97.78%	Highly accurate; no significant difference in accuracy by years.
RN	97.28%	96.36%	97.64%	In all three years, no significant difference was found in E^2 accuracy among program types: ADN, BSN, diploma, and PN.
PN	100%	97.87%	99.12%	
Monitoring	Significant factor ($P = .05$)	Significant factor ($P = .05$)	Not a significant factor	In Years I and II, accuracy significantly improved when test administration was monitored. In Year III, monitoring was not significantly related to E^2 accuracy.
Low-scoring outcomes	No data obtained	Significantly more low-scoring students failed NCLEX than high-scoring students ($P = .001$)	Significantly more low-scoring students failed NCLEX than high-scoring students ($P = .001$)	In both Years II and III (the only two years for which data were obtained), significantly more low-scoring students failed the NCLEX than did high-scoring students.

	Year I	Year II	Year III	Summary of Findings
Remediation **RN students**	No data obtained	Significant factor in decreasing failures ($P =.05$)	No significant difference in NCLEX-RN scores of remediated low-scorers and those not remediated	In Year II, significantly fewer low-scoring students failed the NCLEX-RN when the E^2 was used as a benchmark for remediation. In Year III, this finding was not replicated because of the broad interpretation of the term remediation.

*Accuracy rates were calculated using the most stringent method: total number of predicted to pass

failures divided by the total number predicted to pass.

recognized benchmark for quality patient care within the industry and to identify cost-effective service providers (Rudy et al., 2001). The American Nurses Association (ANA) quickly identified nursing-sensitive patient outcome indicators to demonstrate the role that the nursing profession played in the delivery of quality patient care, and began publishing an annual report card for nursing in 1995 (American Nurses Association, 1995, 1996).

While benchmarking practice is now an accepted practice in most industries within the United States, including the health care industry, it has not been adopted frequently within higher education (Billings et al., 2001). The Scholastic Aptitude Test (SAT) serves as an example of how benchmarking is used in higher education. Every state in the nation adopts a benchmark score in order to select the semifinalists that will be considered for award of a National Merit Scholarship (Hewitt, 2002). The benchmark score is based upon the lowest score among the top percentage of scores (the exact percentage used varies from state to state depending upon the number of high school seniors that reside in each state). Therefore, all students at the junior level that completed the Pre-SAT and achieved a score at or above the named benchmark were named as National Merit Semifinalists. The highest benchmark identified by a state this year was 221 (compared to a perfect score of 240.) In summary, benchmarking has been successfully used in the discipline of education to facilitate objective decision-making about student selection processes based on the results of standardized testing, but there is little published information about the process of establishing educational benchmarks.

Efforts to embrace the practice have begun to appear in the educational literature, but few have specific application to nursing education (Epper, 1999). The few "gold standards" that have been applied in nursing education have been primarily associated with external curriculum evaluation (Wilson, 1999). These evaluation methods were primarily employed by schools of nursing for the purpose of reaffirmation of accreditation by nationally recognized accrediting bodies, such as the NLNAC and CCNE, but have been expanded into the adoption of expansive continuous quality improvement programs implemented to document schools' achievements in meeting high educational standards (Yearwood, Singleton, Feldman, & Colombraro, 2001). The recent trend characterized by society's interest in requiring accountability for educational institutions who have failed to cap spiraling annual cost increases in meeting their stated educational goals has undoubtedly fueled an increased interest in benchmarking in higher education (Epper, 1999; Wilson, 1999). In the midst of declining student enrollment in the nation's schools of nursing, declining NCLEX pass rates, and a shrinking nursing workforce, achievement of a recognized goal in nursing education, success for first-time NCLEX candidates, has emerged as a top priority for stakeholders in nursing education, including students, faculties, administrators, employers, and consumers. However, few authors have described specific benchmarks to identify students at-risk of NCLEX failure, leaving nursing faculties with few guideposts for establishing benchmarks within their programs. The decision-making process faculties use to identify and place students in remediation programs has been primarily based on professional judgment rather than empirical evidence of student competence. A new and growing trend among nursing

62

faculties is to set benchmarks at the end of the curriculum that contain progression-to-graduation requirements based on objective evidence of nursing competency. These policies are designed to identify students in need of remediation prior to graduation and NCLEX candidacy so that NCLEX failure can be avoided.

Published findings regarding the E^2's high degree of accuracy in predicting NCLEX outcomes may have been a catalyst for reports of faculties' recent adoption of specific E^2 scores as benchmarks for progression (Hanks & Lauchner, 1999; Lauchner et al., 1999; Newman et al., 2000; Nibert & Young, 2001). Morrison, Free, & Newman (2002) first described the use of the E^2 as a benchmark for progression. Administrators at seven nursing programs that had implemented policies, which used E^2 scores as a benchmark for progression, were interviewed regarding NCLEX outcomes before and after implementing these policies. Findings indicated that NCLEX-RN pass rates increased by 9 to 41% within two years of implementing the progression policy. However, criteria contained within these policies, including a specific benchmark indicating readiness for graduation, were not described. Benchmarking and other evidence-based approaches to curriculum evaluation and policy development within nursing programs, particularly as these relate to progression-to-graduation and approval-for-NCLEX candidacy issues, lack empirical evidence from research studies. More research findings are needed to guide faculties who wish to adopt benchmarking strategies to demonstrate their adherence to "best practices" in education and attainment of program objectives.

Remediation

While numerous studies have identified effective interventions for students at risk of failing the NCLEX, few have empirically tested the interventions described as effective. In fact, Beeson and Kissling (2001) questioned whether any studies had looked at specific activities that helped at-risk students achieve NCLEX success. Findings of a meta-analysis conducted by Campbell and Dickson (1996) revealed that, of four experimental studies published between 1981 and 1990 that examined the use of interventions, all of these showed significant effectiveness in improving NCLEX performance. Several researchers have reiterated the importance of early identification of at-risk students followed by early intervention to remediate areas of content weakness to effectively avert NCLEX failure (Ashley & O'Neil, 1991; Frierson et al., 1993; Wall et al., 1993). Interventions classified as the most traditional due to their consistent reporting in the literature for over a decade include: enrollment in NCLEX review courses, self-paced study with NCLEX preparatory review books, and completion of simulated NCLEX questions; tutoring by faculty and enrollment in faculty-led or student-led support groups; training in test-taking skills and anxiety and stress-management techniques; assistance by English-as-a-second-language (ESL) instructors for students who are at risk of failing NCLEX due to problems with mastering English language competencies; and use of flash cards or other types of study aids (Ashley & O'Neil, 1991, 1994; Brown, 1987; Frierson et al., 1993; Waterhouse et al., 1993; Wolahan & Wieczorek, 1991). To reduce the rising trend in attrition among culturally diverse students and to reverse the recent decline in NCLEX pass rates, nursing faculties are

increasingly structuring formal and informal support programs intended to bolster academic skills, address nursing knowledge deficits, and facilitate social support systems within the collegiate environment (Burris, 1987; Lockie & Burke, 1999; Marshall, 1989; Memmer, 1991; Rodgers, 1990).

Since the adoption of the CAT version of the NCLEX in 1994, faculties have increasingly indicated a preference for more non-traditional remediation strategies that incorporate the use of computerized review programs for the NCLEX, particularly those that closely simulate the structure of NCLEX test items and required keystrokes to complete the exam (Beeson & Kissling, 2001; Billings et al., 1996; Riner et al., 1997; Ross et al., 1996). Beeson and Kissling (2001) noted that computer-assisted NCLEX-RN review programs were accommodating of students' schedules, an advantage for students experiencing multiple role strain, who felt constant time pressure to fulfill responsibilities unrelated to their nursing studies (Arathuzik & Aber, 1998). Ross et al. (1996) reported favorable indications of reliability and validity for a computerized software program, the "Computer-Assisted Preparation for the NCLEX-RN" published by Williams and Wilkins Electronic Media, Incorporated. Two hundred-thirty undergraduate nursing students' practice examinations administered through the software program were analyzed to determine which students were at-risk of NCLEX failure. The mean score for students failing the NCLEX-RN was 57.25%, whereas the mean score for students passing the NCLEX-RN was 66.81%. This report also included the participating faculty's recommendation for a minimally-acceptable passing score as 65%, which was 12 percentage points higher than the minimally-acceptable passing score recommended by

65

the software publisher. This study was also the only one found that included a specific cut-off score that was used to identify students at risk of failing NCLEX. However, the small sample size and the lack of statistical rigor applied to the reliability and validity analysis reported pose limitations regarding the study's findings. Other researchers supplied subjective reviews of software programs, but did not attempt to quantify the use of these programs by faculties, or indicate the effectiveness of any remediation programs undertaken that incorporated the use of these programs (Billings et al., 1996; Riner et al., 1997).

With regard to the use of the E^2 as a guide for remediation, in the Year II study, Newman et al. (2000) recommended that further research be conducted to evaluate the effectiveness of various types of remediation programs. Contrary to the findings of Year II, in Year III, there was no significant difference in NCLEX-RN success of low-scoring students who were remediated and those who were not remediated. It was postulated that the definition of remediation needed to be more definitive. Remediation was broadly interpreted to be any type of additional study that was based on E^2 findings. The Year IV study was therefore designed to focus more on E^2 implementation strategies than previous studies and to reveal more information about the methods faculties employ to use the E^2 as a remediation guide.

Summary

This literature review encompassed an overview of the decades-long search by nursing educators for predictors of NCLEX success. Studies that focused upon the identification of these predictor variables; their accuracy in predicting NCLEX success

given the changes in this examination over time; and the characteristics of current nursing program applicants that have altered the effectiveness of traditional predictor variables were examined. Additionally, the relatively new development of benchmarking for progression within nursing education programs and the use of computerized NCLEX simulation examinations to guide the selection of remediation strategies for students at risk of NCLEX failure were described. Finally, findings of previous validity studies conducted that focused on the instrument under study, the E^2, were presented. The intent of this review of the literature is to offer background information about the variables of interest in this study. The review offered support for the continued search for consistent predictor variables of NCLEX outcomes and offered background information about the subject of interest in this study.

CHAPTER 3

PROCEDURE FOR COLLECTION AND TREATMENT OF DATA

This descriptive study was conducted to (1) assess the predictive accuracy of the E^2; (2) determine benchmarks established by schools of nursing to determine readiness for graduation and NCLEX candidacy; and (3) assess remediation strategies employed by schools of nursing administering the E^2 to assist students who fail to reach established minimal E^2 benchmarks. A descriptive research approach is used to explain a problem or situation as it currently exists using data that has already been gathered about variables of interest (Dempsey & Dempsey, 1992). This approach is contrasted with the experimental approach, which requires manipulation of variables by the researcher so that predictions can be made about a future occurrence (Dempsey & Dempsey, 1992).

Nursing researchers today have large pools of available nursing data available that are suitable for incorporation into a study designed using a descriptive research approach. Available data sources include computerized databases, reports by statistic-gathering organizations, licensing bureau reports, and the like (Dempsey & Dempsey, 1992). These data offer researchers the opportunity to perform comparisons and contrasts that may not have previously been attempted or reported when the database was constructed. The HESI database, a computerized pool of data consisting of student scores on all HESI examinations, is one such data source. A descriptive, comparative design was used to collect data from the HESI database and complete the analysis of the research

findings. Data regarding E^2 scores during exam administrations in academic year 1999-2000 were obtained from the HESI database. In addition, NCLEX outcomes and responses to a questionnaire were obtained from administrators at schools of nursing that administered the E^2 during the academic year 1999-2000.

Setting

Data were obtained from administrators at RN and PN schools of nursing that administered the E^2 between September 1, 1999 and May 31, 2000 (Year IV). These schools were identified from the HESI database, newly designed in 2000-2001. HESI, founded by a nurse entrepreneur who has published nursing examinations for over a decade using a critical thinking approach, accommodated in the study year approximately 250 schools of nursing in the United States with enrollments of more than 10,000 students annually with computerized custom and standardized examinations. Because of incompatible file formats, some data could not be exported from the database into the questionnaire format. Usable data were available for only 202 of approximately 250 programs that administered the E^2 in 1999-2000. Responses were received from administrators at 158 of the 166 (95.18%) RN programs and 31 of the 36 (86.11%) PN programs which had usable data and were therefore invited to participate in the study.

Population and Sample

A total of 11,988 students took the E^2 during Year IV: 10,546 RN students and 1,442 PN students. Data were exported from the database into the questionnaire format for 6,300 RN students or 59.74% of the total RN student population and 1,035 of the PN students or 71.78% of the total PN population. Respondents consisted of 5,903 (86.81%)

RN students and 897 (13.19%) PN students. Of the 5,903 RN students enrolled in 158

RN programs, 3,459 (58.60%) were enrolled in 92 associate degree (ADN) programs;

2,346 (39.74%) were enrolled in 63 baccalaureate degree (BSN) programs; and 98

(1.66%) were enrolled in 3 diploma programs. The 897 PN students were enrolled in 31

PN programs. Response rates for both the RN and PN programs were considered to be

representative of the 1999-2000 student populations and were comparable in size to the

three previous studies' samples.

Protection of Human Subjects

This study qualified for exempt review as research involving the use of

educational tests, and was approved by the Texas Woman's University Institutional

Review Board—Houston Center in March, 2002. All data obtained from program

administrators were added to the HESI database. Names of the students were not needed

for data entry, and only the total numbers of failures from each group were actually

entered into the HESI database. Additionally, no identifying data regarding schools,

administrators, or students was incorporated into this dissertation or published articles of

the study's findings.

Instrument

The E^2 is a comprehensive, computerized nursing exam that is administered in the

last semester or quarter of a nursing curriculum. It simulates the NCLEX in that it follows

the test blueprint for either the NCLEX-RN or NCLEX-PN developed by the NCSBN

(National Council of State Boards of Nursing, 1999, 2001). Test items for the E^2 are

written using a critical thinking model described by Morrison, Smith, and Britt (1996)

and Morrison and Free (2001) that require application of clinical nursing judgment to determine correct responses. Each version of the E^2 is developed from test banks containing questions written specifically for HESI by a national pool of nurse educators and clinicians.

The HESI Predictability Model (HPM), a proprietary mathematical model, is used to calculate all HESI scores. This calculation does not produce a percentage score. Instead, the HESI score reflects application of the mathematical model to raw scores. The HPM considers the difficulty level of each test item in determining students' performance on the exam. For example, a HESI score of 85 might be a percentage score of 65%, depending on the difficulty level of the test items contained on a particular test or in a particular category of a test. An E^2 report contains a total HESI score as well as scores for clinical specialty areas and sub-topics of these specialty areas. Additionally, HESI scores are provided for five Nursing Process categories, ten NCLEX Client Needs categories (National Council of State Boards of Nursing, 1999, 2001); three National League for Nursing Accrediting Commission (NLNAC) categories (National League for Nursing Accrediting Commission, 1999); and 17 categories described the American Association of Colleges of Nursing (AACN) (American Association of Colleges of Nursing, 1999).

Reliability

Every test returned to HESI for development of an aggregate summary analysis undergoes an item analysis. As a measure of reliability, the Kuder Richardson Formula 20 (KR-20) is calculated for each test analyzed, and the point biserial correlation coefficient is calculated for each test item contained on a test. These data are stored in the

HESI database and used in the calculation of projected reliability for each test administered (Morrison, Adamson, Nibert, & Hsia, 2002). The average KR-20 for the E^2 was 0.75 for the RN group and 0.79 for the PN group, which was comparable to the KR-20 averages reported in previous years' studies.

Validity

Validity of the E^2 is determined by an assessment of content validity, construct validity, and criterion-related validity. Content validity refers to the test items' effectiveness in measuring students' basic nursing knowledge and skills. Expert nurse educators and clinicians established content validity for the E^2 by evaluating the test items' relevance to entry level practice. Construct validity refers to the extent to which a test measures specified traits or attributes at an abstract level. As a comprehensive exit exam, the E^2 measures constructs that are essential to entry level nursing practice as defined by the NCSBN job analysis studies (Chornick & Wendt, 1997) and reflected in the NCLEX test plans (National Council of State Boards of Nursing, 1999, 2001). Criterion-related validity refers to inferences made from analyses of students' E^2 scores used to predict NCLEX success. Annual research studies that correlate E^2 scores with actual NCLEX outcomes offer further evidence of predictive validity (Lauchner, Newman, & Britt, 1999; Newman, Britt, & Lauchner, 2000; Nibert & Young, 2001).

Data Collection

Schools that administered the E^2 received a summary analysis of their aggregate data. One of the reports contained in this summary analysis is a grouping of student's scores by scoring categories. These categories ranged from A, the highest-scoring

72

category, to H, the lowest-scoring category. These scoring categories served as the basis for formulating the scoring intervals that were used for data collection and data analysis. Previous studies examined only high-scoring students, those who scored in categories A and B (90-99.99), and low-scoring students, those who scored in categories G and H ($\leq$69.99). Although the previous studies demonstrated a significant difference in the NCLEX success of high-scoring and low-scoring students, they did not provide data about middle-scoring students. Therefore, in this study, additional scoring intervals were designated to provide more discrimination in the middle-scoring groups where the greatest ambiguity existed regarding the degree of risk for NCLEX failure. The scoring intervals were made up of HESI scores designated as: A/B, scores from 90.00-99.99; C, scores from 85.00-89.99; D, scores from 80.00-84.99; E/F, scores from 70.00-79.99; and G/H, scores $\leq$69.99. A questionnaire, along with a cover letter inviting participation, was mailed to RN and PN program administrators at participating schools of nursing. A list of the school's students who took the E^2 for the first time was included in the mailing. The students' names and E^2 scores were grouped according to scoring intervals. Program administrators were asked how many students in each scoring interval failed the NCLEX. Names of the students were not needed for data entry, only the total numbers of failures within each group. Additional survey data were obtained from school administrators regarding the use of progression and remediation policies based on students' E^2 scores.

Treatment of Data

To answer the first research question regarding the predictive accuracy of the E2, standard statistical methods were used to perform the necessary computations. Predictive

accuracy of the E^2 was calculated using the most stringent method, by examining only the NCLEX outcomes of those who were predicted to pass, which consisted of those scoring in categories A or B. The number of students scoring in category A/B who failed the NCLEX was divided by the total number of students who were predicted to pass, and subtracted from one. Chi squares were calculated to detect differences between expected and observed frequencies among NCLEX outcomes of students scoring at each of the five HESI scoring interval categories, and among students in the different nursing programs scoring at each of the five scoring intervals.

To answer the second research question, descriptive data regarding progression policies were summarized using frequency distributions. Frequency data regarding the use of progression policies were obtained from the questionnaires mailed to participating schools. Administrators were asked if their school had implemented or maintained progression or remediation policies based on E^2 scores. Administrators were also asked to report the minimally-acceptable HESI score required for progression within the curriculum if they required students to attain such a benchmark for progression to graduate or receive approval for NCLEX candidacy.

To answer the third research question, descriptive data were obtained regarding the types of remediation strategies implemented by participating schools that required remediation as a component of their progression policies. The descriptive data supplied by the participating program administrators was used to construct a frequency distribution depicting the most-frequently-used remediation strategies.

CHAPTER 4

ANALYSIS OF DATA

A descriptive, comparative design was used to examine the predictive accuracy of the Health Education Systems Incorporated (HESI) Exit Exam (E^2) for students in all types of nursing programs; identify criteria used by educators to determine students' readiness for graduation as stated in policies of schools administering the E^2; and describe remediation strategies designed to assist students who fail to achieve minimal E^2 benchmarks. Data regarding E^2 scores were obtained from the HESI database. NCLEX outcomes and responses to a questionnaire were obtained from administrators at schools of nursing that administered the E^2 during the academic year 1999-2000. This chapter addresses the description of the Year IV sample and the findings generated from data analysis.

Description of the Sample

Data were obtained from administrators or their designees at RN and PN schools of nursing that administered the E^2 in Year IV. These schools were identified from the HESI database, newly designed in 2000-2001. Because of incompatible file formats, some data could not be exported from the database to the questionnaire format. Consequently, the population for this study was limited to only those schools whose data could be successfully exported.

Respondents consisted of 5,903 (86.81%) RN students and 897 (13.19%) PN students. Of the 5,903 RN students enrolled in 158 RN programs, 3,459 (58.60%) were enrolled in 92 associate degree (ADN) programs; 2,346 (39.74%) were enrolled in 63 baccalaureate degree (BSN) programs; and 98 (1.66%) were enrolled in 3 diploma programs. Table 4 shows the breakdown of the RN sample by type of program. The 897 PN students were enrolled in 31 PN programs.

Findings

Predictive Accuracy

The first research question addressed the accuracy of the E^2 for Year IV for both the RN and PN groups. In Year IV, a total of 6,800 nursing students, 5,903 RN and 897 PN students, comprised the study sample. A total of 1,303 (22.07%) ADN students, 726 (12.30%) BSN students, and 30 (0.51%) diploma students, as well as 341 (38.02%) PN students received HESI scores in the A/B category (90.00-99.99), indicating that they were predicted to pass the NCLEX-RN or NCLEX-PN without additional preparation. Of the 2,059 RN students who scored in the A/B category, 35 (1.70%) failed the licensure exam. In the ADN group, 21 (1.61%) of the 1,303 students who scored in the A/B category failed, 14 (1.93%) of the 726 BSN students who scored in the A/B category failed, and none (0%) of the 30 diploma students scoring in the A/B interval failed. Of the 341 PN students who scored in the A/B category, 2 (0.59%) failed the NCLEX-PN.

The predictive accuracy of the E^2 in Year IV was 98.30% for RN students, 99.41% for PN students, and 98.46% for all students. Results of a chi square goodness-

Table 4.

RN Sample Breakdown by Type of Program

Description	ADN	BSN	Diploma	Total
Number	3,459	2,346	98	5,903
Percentage	58.60%	39.74%	1.66%	100%

of-fit revealed that the Year IV predictive accuracy for the all students (98.46%) was not significantly different χ^2 (9, $N = 6,300$) = .119, $p = .001$) from that of Year I (97.41%), Year II (96.49%), or Year III (97.78%). Thus, as in the three previous years of study, the predictive accuracy of the E^2 was not significantly different among the years of study. Also, as was the case with the three previous years of study, there was no significant difference in the predictive accuracy by types of programs examined: ADN, BSN, diploma, or PN programs.

Predictive Accuracy by Scoring Intervals

Previous studies examined only the outcomes of the high-scoring and low-scoring E^2 students, and a similar examination of RN students' scores in Year IV found that significantly more χ^2 (1, $N = 5903$) = 571.401, $p = .001$ of the low-scoring students failed NCLEX than did the high-scoring students. However, Year IV examined not only high-scoring and low-scoring students, but also compared NCLEX outcomes for the five HESI scoring intervals. A chi square analysis of the students' scores for each of the five scoring

77

intervals revealed significant differences χ^2 (4, N = 5903) = 1045.630, p = .001 among

scoring intervals of RN students' scores as well as significant differences χ^2 (4, N = 897)

= 301.057, p = .001 among scoring intervals of PN students' scores.

Analysis of scoring interval data indicated that NCLEX failures increased as the scoring

interval decreased. Of the 2,059 RN students who scored in the A/B category, 35 (1.70%)

failed the licensing exam; of the 1,014 students who scored in the C category, 60 (5.92%)

failed; of the 980 students who scored in the D category, 106 (10.82%) failed, of the

1,324 students scoring in the E/F category, 314 (23.72%) failed, and of the 526 students

scoring in the G/H category, 264 (50.19%) failed. Figure 2 illustrates the patterns of

NCLEX-RN success and failure associated with each of the different scoring categories.

Of the 341 PN students who scored in the A/B category, 2 (0.59%) failed the licensing

exam; of the 192 students who scored in the C category, 10 (5.21%) failed; of the 144

students who scored in the D category, 20 (13.89%) failed, of the 167 students scoring in

the E/F category, 75 (44.91%) failed, and of the 53 students scoring in the G/H category,

38 (71.70%) failed. Figure 3 illustrates the patterns of NCLEX-PN success and failure

associated with each of the different scoring categories.

Students' outcomes were compared by program type for each of the scoring

intervals, e.g., A/B interval scores of ADN students were compared with A/B interval

scores of BSN students. Scores at all intervals were compared for each of the four

programs until an analysis of all possible combinations of programs was completed. No

significant differences were found in the predictive accuracy of the E^2 among programs
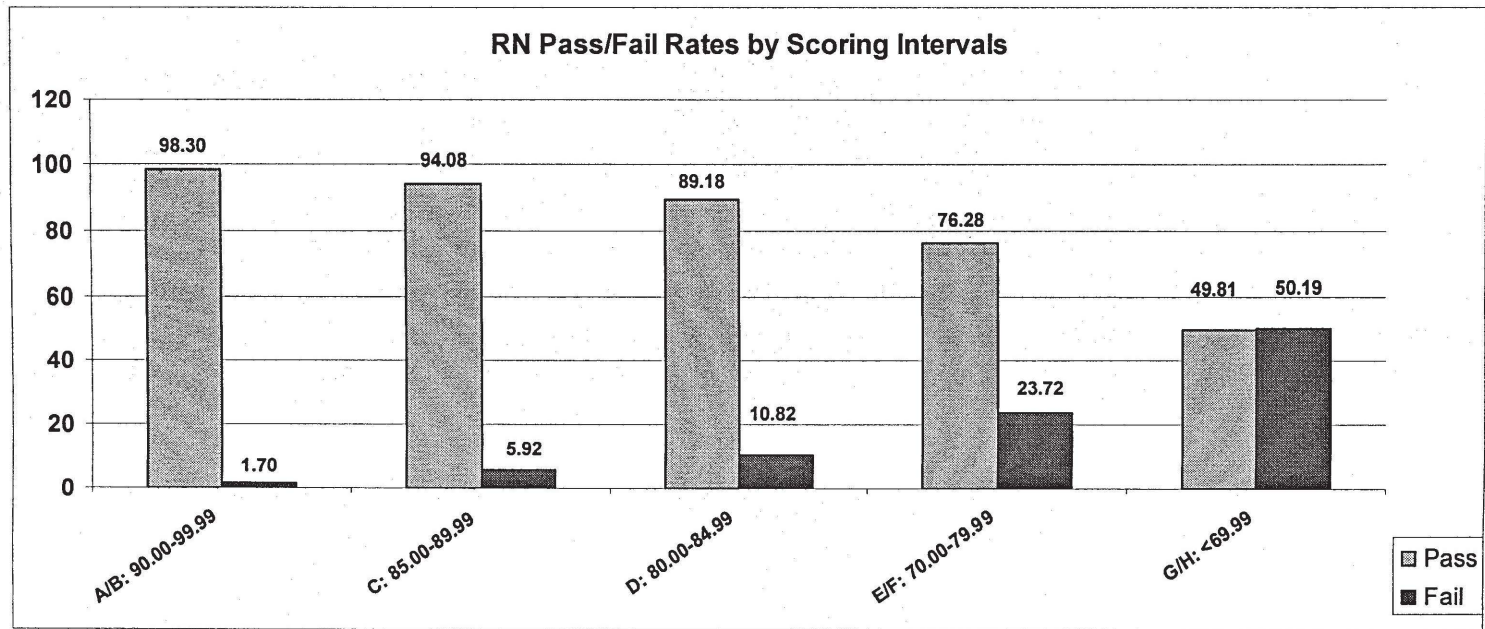
RN Pass/Fail Rates by Scoring Intervals

A/B: 90.00-99.99 — Pass 98.30, Fail 1.70
C: 85.00-89.99 — Pass 94.08, Fail 5.92
D: 80.00-84.99 — Pass 89.18, Fail 10.82
E/F: 70.00-79.99 — Pass 76.28, Fail 23.72
G/H: <69.99 — Pass 49.81, Fail 50.19

Figure 2. NCLEX-RN Pass/Fail Rates by E^2 Scoring Intervals

(ADN, BSN, diploma, or PN) for students scoring at like levels in each of the five scoring intervals. In other words, significant differences in NCLEX pass rates were noted for each scoring interval, but no significant difference was found among different programs when they were compared at like intervals. However, when aggregate data regarding NCLEX outcomes for all RN programs were compared with aggregate data for all PN programs at each of the five scoring intervals, a significant difference was found between RN and PN students in the lowest-scoring categories only, the E/F (70.00-79.99) and G/H ($\leq$69.99) categories. Significantly more RN students scoring in the E/F χ^2 (1, N = 1491) = 34.545, p = .001 and G/H χ^2 (1, N = 579) = 8.926, p = .003 categories passed the NCLEX as compared to PN students scoring in these two categories.

Use of the E^2 as a Benchmark for Progression and Remediation

The second research question addressed the E^2 score criteria that nursing schools used to determine readiness for graduation. NCLEX results and frequency data regarding remediation and progression policies were obtained from the questionnaires mailed to participating schools. Administrators were asked if their school had implemented a progression or remediation policy based on E^2 scores. Of the 158 participating RN programs, 149 (94.30%) indicated that they used E^2 scores as a benchmark for remediation, and 45 (30.20%) of these schools tied progression to a minimally acceptable E^2 score in order for students to graduate or to take the licensing exam. Of the 31 participating PN programs, 5 (16.13%) indicated that they used E^2 scores as a benchmark
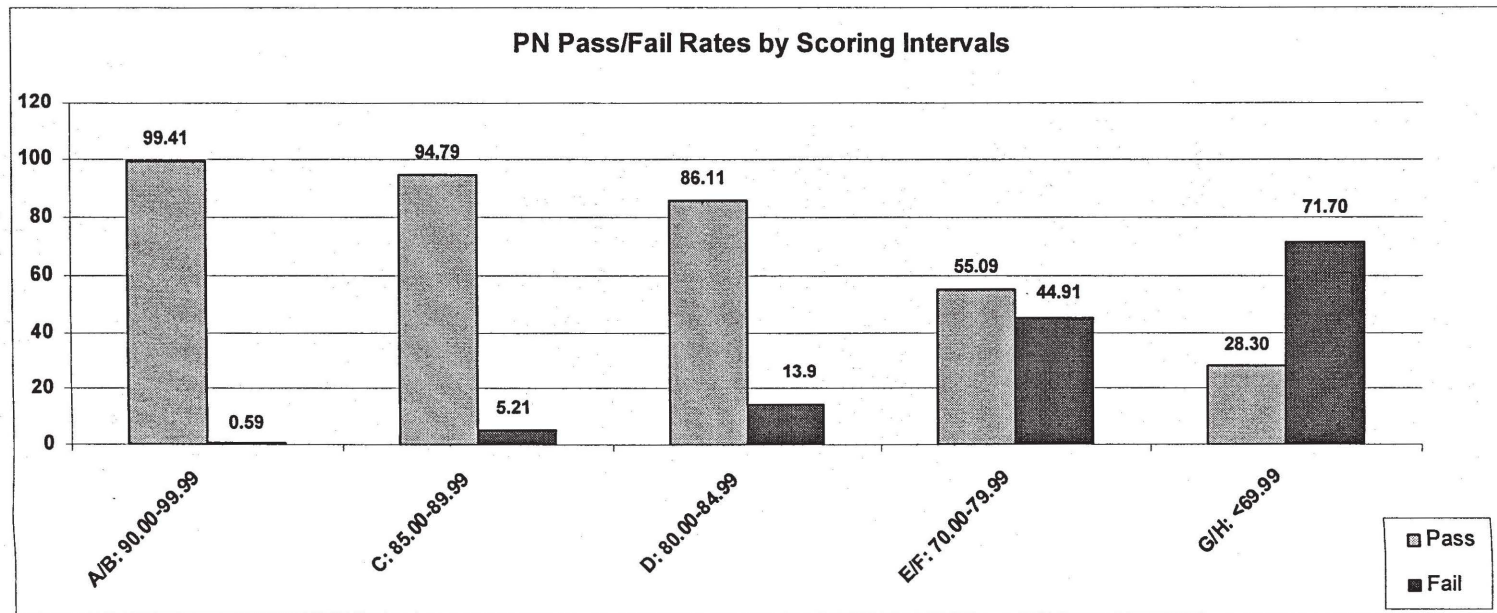
Figure 3. NCLEX-PN Pass/Fail Rates by E² Scoring Intervals

for remediation, and these same 5 schools tied progression to a minimally acceptable E^2 score in order for students to graduate or to take the licensing exam.

Of the 45 RN programs that implemented new or existing progression policies, 35 (77.78%) submitted either complete or partial progression policy statements from their schools. Based on a review of the submitted policies, one or more consequences were most often cited for students who did not achieve the benchmark E^2 score designated by the school, including the assignment of an incomplete or failing grade in the capstone course (12 or 34.29%); denial of eligibility for graduation (18 or 51.43%); and/or withholding of approval for NCLEX candidacy (5 or 14.29%). These findings are summarized in Figure 4.

Additional data were obtained from HESI database regarding enforcement of mandatory re-testing for students who failed to meet established benchmarks at the 45 schools that had adopted progression policies. In 1999-2000, 36 (80.00%) of these schools enforced mandatory re-testing using a different version of the E^2. One year later, in the 2000-2001 academic year, four more of these programs required re-testing for students who failed to attain the stated benchmarks. Therefore, 40 (88.89%) of the participating programs enforced mandatory re-testing with a different E^2 version for students that failed to achieve their schools' specified E^2 benchmarks.

Of the 35 progression policies submitted, 20 (57.14%) specified the number of re-tests with a different version of the E^2 that were permitted: 7 (20.00%) permitted one re-testing; 9 (25.71%) permitted two re-testings; and 4 (11.43%) permitted an unlimited
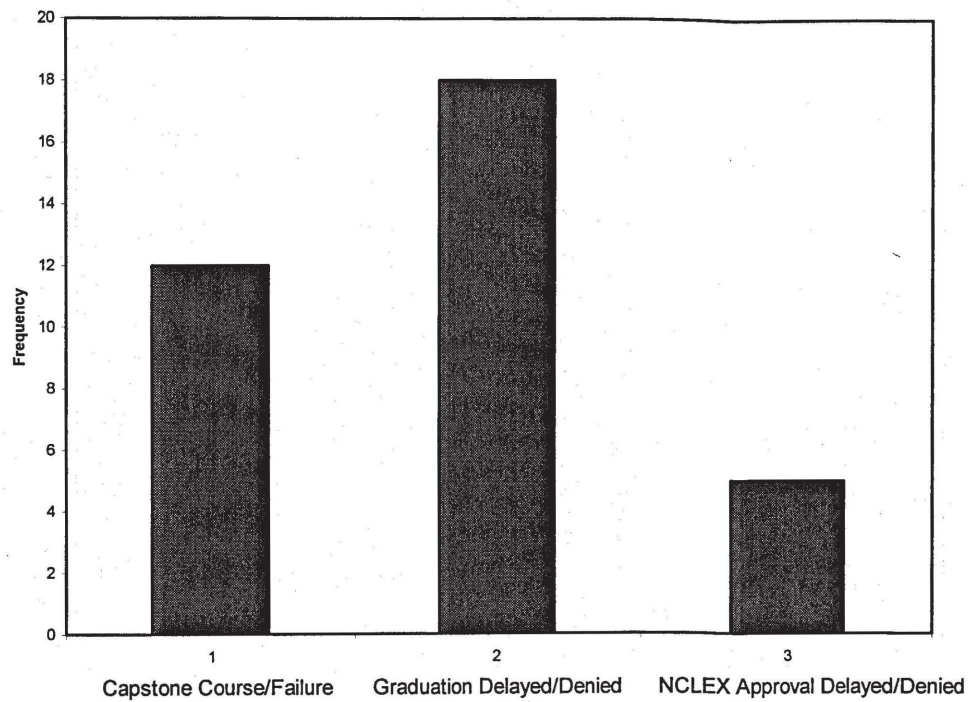
Figure 4.   Progression Policies: Consequences of Benchmark Failure ($N = 35$)

number of re-testings. Program administrators also reported that typically their schools covered the cost of the first E^2, which was either attributed to a school fee or a fixed program cost, but that students who re-tested were required to pay for all repeated exam administrations.

Benchmarking: E^2 Score Needed to Progress

Program administrators were asked if a minimal HESI score on the E^2 was required for progression to graduation or NCLEX candidacy. Of the 158 participating RN programs, 148 (93.67%) administrators responded to this question, and 45 (30.41%) of these respondents indicated that they had established a policy that designated benchmark E^2 HESI score for progression. These benchmark scores ranged from 77 to 90. Most, 36 (80.00%), of the programs reported using an E^2 score of 85 as the benchmark for progression. Seven (15.56%) RN schools adopted E^2 benchmarks scores above 85: three selected 90, two selected 88, and two used 87. Two (4.44%) schools chose scores below 85: one selected 80 and the other used 77.

Of the 5 PN schools that reported using E^2 scores as a benchmark for remediation, 2 (40.00%) reported using an E^2 score below 85 (both used a score of 75), while only one (20.00%) used an E^2 score higher than 85 (the score used was 86).

Remediation Strategies

Program administrators were asked if remediation was required as a function of the school's progression policy. Of the 158 participating RN programs, 149 (94.30%)

submitted responses to this question. Most program administrators (107 or 71.81%) reported that remediation was not required. However, these administrators described *optional* remediation plans available to students who did not meet the minimal E^2 requirement for progression. The optional remediation strategies used were the same as the required strategies reported; however, no consequences were involved if students failed to complete the remediation activities. Of the 42 (28.19%) remaining RN schools that required remediation, the strategies implemented were: a specially designed remediation course (22 or 52.38%); completion of software programs offering computer assisted instruction (10 or 23.81%); comprehensive review guided by NCLEX preparation books (4 or 9.52%); mandatory tutoring sessions with faculty (4 or 9.52%); and completion of an NCLEX simulation exam (one or 2.38%). Figure 5 describes the distribution of remediation strategies most frequently used by participating RN schools. Due to the extremely small number of PN schools (a total of 5) that enforced policies requiring attainment of a specific E^2 benchmark score, attempts to summarize remediation strategies were abandoned, as the findings on such a small sample were not meaningful.

Summary of the Findings

The E^2 demonstrated a high degree of accuracy in predicting NCLEX success (98.46%) for students who took the exam during academic year 1999-2000. The predictive accuracy of the E^2 was not significantly different among the programs tested (ADN, BSN, diploma, or PN). Additionally, the predictive accuracy of the E^2 in Year IV
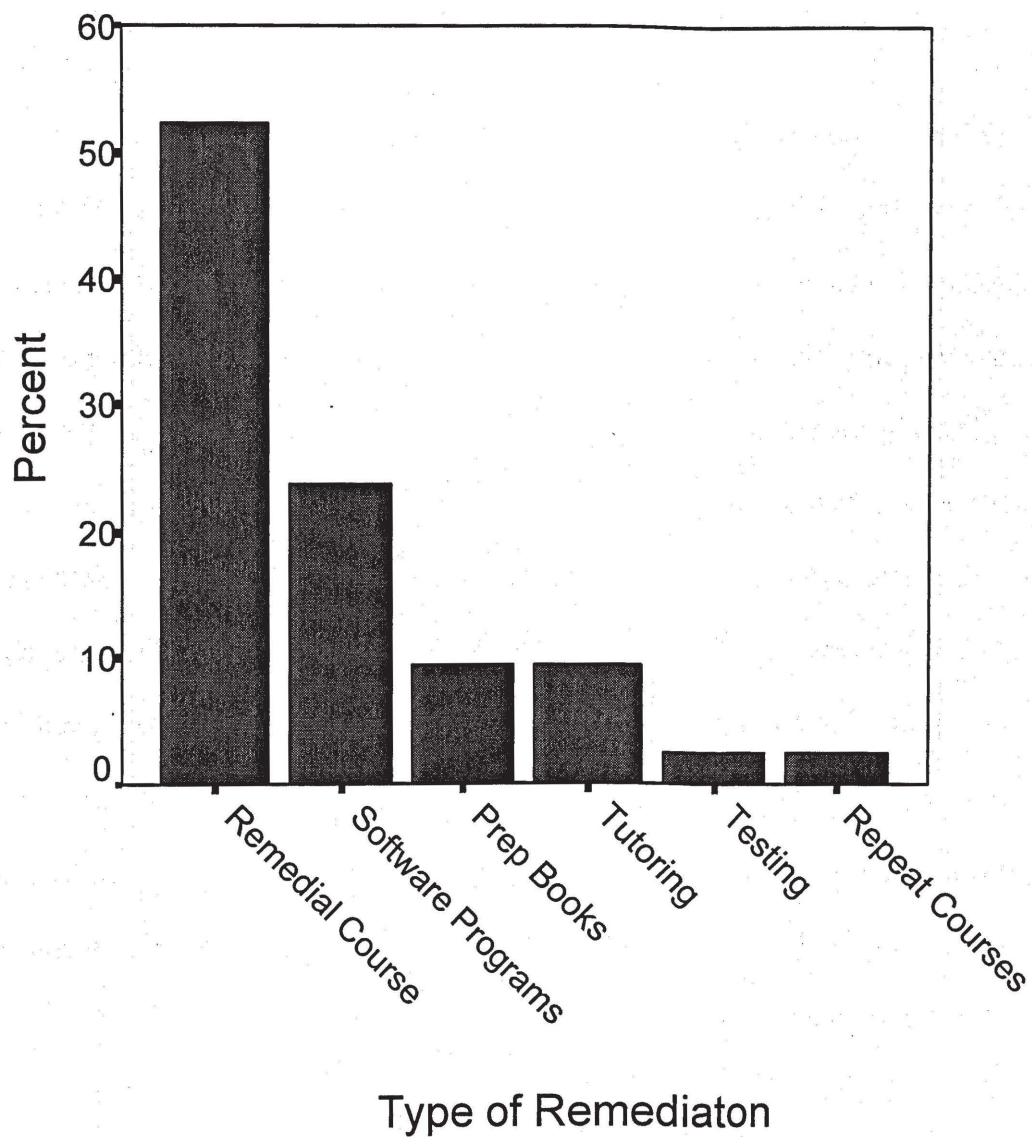
Figure 5. Percentages of RN Programs using Various Types of Remediation

($N = 42$)

was not significantly different from the predictive accuracy of the E^2 in the three prior years of study. An analysis of students' E^2 performance by specific scoring intervals revealed a consistent pattern. The percentage of students who failed the NCLEX significantly increased with each successive drop in scoring interval, creating a step-wise pattern of progressively higher percentages of subjects failing the NCLEX-RN. Regarding the adoption or maintenance of school policies stipulating mandatory attainment of a specific E^2 benchmark score as a requirement for progression to graduation, approximately one-third (28.19%) of the RN schools and one-sixth (16.13%) of the PN schools reported having implemented such a policy. Consequences of failure to meet the E^2 benchmark score included: the assignment of an incomplete or failing grade in the capstone course; denial of eligibility for graduation; and/or withholding of approval for NCLEX candidacy. These benchmark scores ranged from 77 to 90. Most of the RN programs reported using an E^2 score of 85 as the benchmark for progression. Seven RN schools adopted E^2 benchmarks scores above 85: three selected 90, two selected 88, and two used 87. Two schools chose scores below 85: one selected 80 and the other used 77. Of the PN schools that reported using E^2 scores as a benchmark for remediation, 2 reported using an E^2 score below 85 (both used a score of 75), while only one used an E^2 score higher than 85 (the score used was 86).

Few PN schools required remediation for students who failed to reach established benchmarks, thus remediation strategies were not reported in detail. However, of the RN schools that required remediation, the strategies implemented were: a specially designed

87

remediation course; completion of software programs offering computer assisted

instruction; comprehensive review guided by NCLEX preparation books; mandatory

tutoring sessions with faculty; and completion of an NCLEX simulation exam.

CHAPTER 5

SUMMARY OF THE STUDY

The purpose of this study was to determine the predictive accuracy of the Health Education Systems, Inc (HESI) Exit Exam (E^2) for the fourth consecutive year, examine the degree of risk for failing the NCLEX associated with various E^2 scoring intervals, and describe frequency data obtained regarding use of the E^2 as a benchmark for progression and remediation. A theoretical framework incorporating elements of critical thinking theory and test (psychometric) theory was used as a basis for this fourth HESI validation study. This chapter contains a summary of the investigation, a discussion of findings, investigational conclusions, implications, and recommendations for further study.

Summary

A descriptive, comparative design was used to examine the data provided by schools of nursing regarding students' NCLEX outcomes in the academic year 1999-2000. A total of 11,988 students took the E^2 during Year IV, 10,546 RN students and 1,442 PN students. Data were exported from the database into the questionnaire format for 6,300 RN students or 59.74% of the total RN student population and 1,035 of the PN students or 71.78% of the total PN population. A questionnaire, along with a cover letter inviting participation, was mailed to RN and PN program administrators at participating schools of nursing. A list of the school's students who took the E^2 for the first time was included in the mailing. The students' names and E^2 scores were grouped according to

scoring intervals. Program administrators were asked how many students in each scoring interval failed the NCLEX. Additional survey data were obtained from school administrators regarding the use of progression and remediation policies based on students' E^2 scores. Responses were received from administrators at 158 of the 166 (95.18%) RN programs and 31 of the 36 (86.11%) PN programs.

Standard statistical methods were used to compute the accuracy of the E^2 in predicting NCLEX-RN and NCLEX-PN success. The number of students scoring in the highest scoring interval on the E^2 who were predicted to pass, but failed the NCLEX, was divided by the total number of students in the highest scoring interval who were predicted to pass, and actually did pass the NCLEX. The number obtained from this division was then subtracted from one to allow for conversion to a percentage figure that represented the exam's predictive accuracy with regard to NCLEX outcome. Chi squares were calculated to detect differences between expected and observed frequencies among NCLEX outcomes of students scoring at each of the five HESI scoring interval categories, and among students in the different nursing programs scoring at each of the five scoring intervals.

Descriptive data regarding progression and remediation policies were summarized using frequency distributions. Approximately one-third (45 or 30.20%) of the 149 responding RN programs indicated that they had implemented a policy that used HESI scores on the E^2 as a benchmark for progression, while only one-sixth (5 or 16.13%) of the 31 responding PN programs had implemented this type of policy. Most of the respondents that had implemented such policy (36 or 80.00%) adopted a HESI score of

90

85 as the minimally-acceptable E^2 benchmark for progression. Although the strategies used for remediation were typical of those described in previous research studies, the use of a progression policy that required students to finish remediation activities as a condition of graduation or NCLEX candidacy for students who failed to attain specified E^2 benchmarks was identified as a new trend.

Discussion of the Findings

Predictive Accuracy of the E^2 Regarding NCLEX Outcomes

The E^2 continued to demonstrate a high degree of accuracy in predicting NCLEX success (98.46%). For the fourth consecutive year, the predictive accuracy of the E^2 was not significantly different among the programs tested (ADN, BSN, diploma, or PN), indicating that the E^2 is an effective predictor of NCLEX success for all types of nursing programs. Additionally, the predictive accuracy of the E^2 in Year IV was not significantly different from the predictive accuracy of the E^2 in the three prior years of study, indicating that, based on the aggregate data collected from 19,554 subjects over four consecutive years, the E^2 is a highly accurate predictor of NCLEX success.

In Year IV, an analysis of students' E^2 performance by specific scoring intervals revealed a consistent pattern. The percentage of students who failed the NCLEX significantly increased with each successive drop in scoring interval, creating a step-wise pattern of progressively higher percentages of subjects failing the NCLEX-RN. The pattern exhibited by the PN students was similar to that of the RN students, but the degree of risk for failure of the NCLEX-PN was more pronounced in the lowest two scoring intervals. This pattern could be a reflection of the lack of remediation reported by

91

the PN schools since only 16.13% of these schools required remediation for those students who obtained low E^2 scores.

The fact that there was no significant difference in scoring interval findings among RN program types indicates that inferences regarding risk for NCLEX-RN failure are the same regardless of program type. However, the finding that significantly more RN students who scored in the two lowest scoring intervals passed the NCLEX than did PN students who scored in these same categories is likely attributable to the fact that PN students usually take the E^2 just before completion of their programs and do not have the opportunity to remediate. The percentage of failures at each scoring interval provides the data necessary to make inferences regarding the degree of risk for NCLEX failure associated with each scoring interval.

The results of the predictive validity analysis of the E^2 regarding NCLEX success support the use of standardized testing in the later half of the student's senior year in nursing school to identify those in need of remediation prior to NCLEX candidacy. Although several comprehensive nursing examinations have been reported to have a moderate to high ability to predict NCLEX success, many of the tests studied require paper and pencil administration, which does not provide students with practice using the NCLEX CAT format (Alexander & Brophy, 1997; Barkley, Rhodes, & Dufour, 1998; Breyer, 1984; Endres, 1997; Ross, Nice, May, & Billings, 1996; Schmidt, 2000). Additionally, according to Beeman and Waterhouse (2001), the predictive validity study of the E^2 published by Lauchner, Newman, and Britt (1999) was the only study of predictor variables that was conducted post-1994, the year the NCSBN adopted the

92

NCLEX CAT format. Therefore, nurse educators searching for a reliable predictor variable, such as student outcomes on a standardized examination that simulates the NCLEX, and who wish to identify students in need or remediation prior to NCLEX, find support for using the E^2 for these purposes as a result of the Year IV analysis.

Progression Policies

Results of this study confirm that benchmarking for progression, a new trend in nursing education, has emerged in all types of nursing programs leading to RN and PN licensure. Approximately one-third (45 or 30.20%) of the 149 responding RN programs indicated that they had implemented a policy that used HESI scores on the E^2 as a benchmark for progression, while only one-sixth (5 or 16.13%) of the 31 responding PN programs had implemented this type of policy. However, since these data were collected, HESI reports many more schools have implemented such policies or are considering implementing such policies.

Between 1999 and 2001, of the participating program administrators that had implemented progression policies re-tested students using a different version of the E^2 following the required remediation. This finding indicates that programs are not only increasingly adopting progression policies that require attainment of a specific E^2 score, but that they are also evaluating the effectiveness of the remediation strategies implemented through re-testing.

Benchmarking: E^2 Score Needed to Progress

The overwhelming majority (40 or 88.89%) of the administrators who adopted a specific E^2 benchmark for progression chose a HESI score of 85 as the minimally-

93

acceptable score. Findings indicated that 98.30% of the RN students who achieved HESI scores between 90.00 and 99.99 and 94.08% of the RN students who achieved HESI scores between 85.00 and 89.99 were successful in passing the NCLEX on their first attempt. Those findings were based on scores obtained from students in associate degree, baccalaureate degree, and diploma nursing programs who took the E^2 for the first time.

While analysis of these data indicated that most schools with established progression policies required a minimal HESI score of 85 for progression, faculties considering adoption of such a policy must carefully examine the characteristics of their programs that could influence their decisions about the use of a specific benchmark. The first consideration must be the number of times the student re-tests with different versions of the E^2 before achieving the minimally-acceptable benchmark score. Remediation efforts are likely to render students who require multiple attempts to achieve the desired E^2 benchmark better prepared for NCLEX than they would have been without such remediation and re-testing. However, they are more likely to be at greater risk for NCLEX failure than those who attain the benchmark score on their first attempt.

Another factor to be considered by faculties contemplating adoption of a minimal E^2 score for progression is the size of the graduating class. Programs graduating very few students can tolerate less risk of NCLEX failure. In small programs, annual pass rates are adversely affected by even one failure, thereby placing these schools at risk of losing accreditation or receiving greater scrutiny by the state board of nursing. Therefore, faculties with small graduating classes may choose a more conservative approach and select a higher E^2 benchmark than schools with larger numbers of graduates.

94

Historically, according to Epper (1999), higher education has not embraced benchmarking as readily as business and industry. However, with increasing demands for accountability by consumers of higher education, the practice of incorporating benchmarking methods to faciliate external curriculum evaluation has gained greater acceptance on university campuses. Schools of nursing have increasingly adopted benchmarking practices that incorporate the use of the E^2 or other standardized examinations to facilitate objective decision-making about student readiness for graduation and NCLEX candidacy. Benchmarking for progression is not only a new trend in nursing education, but one that may signify a turning point in higher education. High stakes testing for the purpose of benchmarking in higher education is increasingly prevalent on many college and university campuses. This trend is likely to continue given the public's increasing demand for accountability in the academic arena.

Remediation Strategies

Remediation interventions described by the participating administrators were typical of those previously identified in the literature (Arathuzik & Aber, 1998; Frierson, Malone, & Shelton, 1993; Memmer, 1991; Messmer, Abelleira, & Erb, 1995; Rodgers, 1990; Symes, Tart, Travis, & Toombs, 2002; Vance, 1997; Wolahan & Wieczorek, 1991). However, one important new trend was confirmed: schools have begun to tie the completion of remediation to the approval for graduation or NCLEX candidacy for students who fail to attain specified E^2 benchmarks. This trend validates the conclusions reported by Morrison, Free, and Newman (2002) which stated that progression alone was

95

enough to motivate students to study so that they would meet designated E^2 benchmarks, ultimately becoming prepared to pass the NCLEX.

A meta-analysis conducted by Campbell and Dickson (1996) revealed that there were few experimental studies published that examined the effectiveness of educational intervention, including remediation programs, that showed significant improvement in NCLEX outcomes. However, with the identification of a new trend tying progression to the completion of remediation to assist students with achievement of program benchmarks, studies that focus on the effectiveness of remediation as a requirement for progression would offer educators valuable information about the usefulness of specific remediation strategies.

Conclusions and Implications

Findings from this investigation support the following conclusions:

1. The E^2 is highly accurate in predicting NCLEX success in both RN and PN programs.

2. The percentage of students who failed the NCLEX significantly increased with each successive drop in scoring interval, creating a step-wise pattern of progressively higher percentages of subjects failing the NCLEX-RN.

3. The most common E^2 score benchmark set by Schools of Nursing is a HESI score of 85.

4. Schools of nursing have begun to tie the completion of remediation interventions, which are most often accomplished using conventional pedagogical methods, to

96

the approval for graduation or NCLEX candidacy for students who fail to attain specified E^2 benchmarks.

Several implications can be derived from this study. Findings regarding the validity of the E^2 will assist nurse educators in explaining to students their risks for NCLEX failure and provide the empirical data that may be useful in convincing students of the necessity to remediate prior to taking the NCLEX. The E^2 has consistently been identified as a highly accurate predictor of NCLEX outcomes. Thus, many nursing faculties have implemented policies that use E^2 scores as progression and remediation benchmarks. When establishing such policies, it is the school's responsibility to designate the required E^2 score for progression. Results of this study regarding the degree of risk associated with various scoring intervals provide the evidenced-based support that faculties need when determining specific E^2 scores to use as benchmarks for progression and remediation. Additionally, these findings may assist faculties who are debating the value of adopting a progression and remediation policy.

Assisting students to complete the nursing curriculum and helping new graduates become successful first-time NCLEX-RN candidates have always been high priorities for nursing faculties. However, the recruitment of more diverse populations into nursing has made these goals more difficult to achieve. Such recruitment has both positive and negative effects. Undoubtedly, academically-at-risk students are provided educational opportunities that might not have otherwise been available to them. However, nursing faculties may also face higher attrition rates and decreasing NCLEX pass rates as a result of these efforts to increase enrollment in nursing programs. The findings of this study

indicate that the use of E^2 scores as benchmarks for progression was effective in providing a guide for remediation, which enabled faculties to better assist students to complete the nursing curriculum and become successful NCLEX candidates.

Recommendations for Further Study

Suggestions for future research are as follows:

1. Although the predictive accuracy of the E^2 has been well established in the consecutive four years of study, periodic evaluations of the exam's validity in predicting NCLEX success should be conducted, with the designs strengthened as much as possible to control these intervening variables.

2. Findings indicated that schools of nursing are implementing policies that use E^2 scores as a benchmark for progression and remediation. Further study is needed to determine the direction of this new trend among all nursing programs and to evaluate the effectiveness of these policies in reducing the risk of NCLEX failure.

3. A quasi-experimental approach should be used to study the effectiveness of specific remediation strategies adopted by individual nursing programs. Such a research design would require the use of a smaller sample of students who consent to reporting their outcomes on both the E^2 and the NCLEX, and it should also include detailed descriptions of the types of remediation undertaken by low-scoring students.

4. Future research should also examine the effectiveness of such strategies with minority and English-as-a-second-language (ESL) students who have traditionally been plagued with less success on the NCLEX than their white, American-born

counterparts (Arathuzik & Aber, 1998; Endres, 1997; Fearing, 1997; Frierson et al., 1993; Tucker, 1999).

5. Further research is needed to establish the specific degree of risk for NCLEX failure associated with the number of times students are allowed to re-test with different versions of the E^2.

6. The NCLEX success of these E^2 re-testers should be examined, with comparisons made between those who are remediated only and those who are remediated and retested using a different version of the E^2.

7. Research is needed to establish the degree of relationship among class size, E^2 benchmarks, and NCLEX outcomes for students in all types of nursing programs.

References

Alexander, J., & Brophy, G. (1997). A five-year study of graduates' performance on

    NCLEX-RN. *Journal of Nursing Education, 36*(9), 443-445.

Allen, M., Nunley, J., & Scott-Warner, M. (1988). Recruitment and retention of black

    students in baccalaureate nursing programs. *Journal of Nursing Education, 27*(3),

    107-116.

American Association of Colleges of Nursing. (1999). *CCNE accreditation standards.*

    Retrieved July, 2000 from AACN-CCNE Web site:

    http://www.aacn.nche.edu/Accreditation/standrds.htm.

American Nurses Association. (1995). *Nursing report card for acute care.* Washington,

    DC: American Nurses Publishing.

American Nurses Association. (1996). *Nursing quality indicators, definitions and*

    implications. Washington, DC: American Nurses Publishing.

Arathuzik, D., & Aber, C. (1998). Factors associated with National Council Licensure

    Examination-Registered Nurse success. *Journal of Professional Nursing, 14*(2),

    119-126.

Ashley, J., & O'Neil, J. (1991). The effectiveness of an intervention to promote

    successful performance on NCLEX-RN for baccalaureate students at risk for

    failure. *Journal of Nursing Education, 80*(8), 360-366.

Ashley, J., & O'Neil, J. (1994). Study groups: Are they effective in preparing students

    for NCLEX-RN? *Journal of Nursing Education, 33*(8), 357-364.

Barbee, E. (2001). Our dismal progress: The recruitment of non-whites into nursing. *Journal of Nursing Education, 30*(6), 243-244.

Barkley, T., Rhodes, R., & Dufour, C. (1998). Predictors of success on the NCLEX. *Nursing and Health Care Perspectives, 19*(3), 132-137.

Beeman, P., & Waterhouse, J. (2001). NCLEX-RN performance: Predicting success on the computerized examination. *Journal of Professional Nursing, 17*(4), 158-165.

Beeson, S., & Kissling, G. (2001). Predicting success for baccalaureate graduates on the NCLEX-RN. *Journal of Professional Nursing, 17*(3), 121-127.

Billings, D., Connors, H., & Skiba, D. (2001). Benchmarking best practices in web-based nursing courses. *Advances in Nursing Science, 23*(3), 41-52.

Billings, D., Hodson-Carlton, K., Kirkpatrick, J., Aaltonen, P., Dillard, N., Richardson, V., Siktberg, L., & Vinten, S. (1996). Computerized NCLEX-RN preparation programs: A comparative review. *Computers in Nursing, 14*(5), 272-286.

Birx, E. (1993). Critical thinking and theory-based practice. *Holistic Nursing Practice, 7*(3), 21-27.

Bloom, B. (1956). *Taxonomy of educational objectives*. New York: David McKay Co.

Brown, M. (1987). The effects of a support group on student attrition due to academic failure. *Journal of Nursing Education, 26*(8), 324-327.

Buerhaus, P., Staiger, D., & Auerbach, D. (2000). Implications of an aging Registered Nurse workforce. *Journal of the American Medical Association, 283*(22), 2948-2954.

Burris, B. (1987). Reaching educationally disadvantaged students. *American Journal of Nursing, 87*(10), 1359-1360.

Campbell, A., & Dickson, C. (1996). Predicting student success: a ten-year review using integrative review and meta-analysis. *Journal of Professional Nursing, 12*(1), 47-59.

Carpenter, D., & Bailey, P. (1999). Predicting success on the Registered Nurse Licensure Examination--1999 Update. In K. Stevens & V. Cassidy (Eds.), *Evidence-based teaching: Current research in nursing education* (pp. 135-170). Sudbury, MA: Jones and Bartlett Publishers and NLN Press.

Cascio, R., Campbell, D., Sandor, M., Rains, A., & Clark, M. (1995). Enhancing critical thinking skills: Faculty-student partnerships in community health nursing. *Nurse Educator, 20*(2), 38-43.

Chornick, N., & Wendt, A. (1997). NCLEX-RN: From job analysis study to examination. *Journal of Nursing Education, 36*(8), 378-382.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth-Thomson.

Dempsey, P., & Dempsey, A. (1992). *Nursing research with basic statistical applications* (Third ed.). Boston: Jones and Bartlett Publishers.

Dexter, P., Applegate, M., Backer, J., Claytor, K., Keffer, J., Norton, B., & Ross, B. (1997). A proposed framework for teaching and evaluating critical thinking in nursing. *Journal of Professional Nursing, 13*(3), 160-167.

Endres, D. (1997). A comparison of predictors of success on NCLEX-RN for African

    American, Foreign-born, and White baccalaureate graduates. *Journal of Nursing*

    Education, 36(8), 365-371.

Epper, R. (1999). Applying benchmarking to higher education: some lessons from

    experience. *Change X, 6*, 24-31.

Facione, N. (1995). *Critical thinking and clinical judgment: Goals 2000 for nursing*

    science. Paper presented at the Annual Meeting of the Western Institute for

    Nursing, San Diego, CA.

Facione, N. (1999a). *California Academic Press Web site.* California Academic Press,

    Millbrae, CA. Retrieved September, 1999 from: http://www.calapress.com.

Facione, N. (1999b). *Meta-Analysis Project Completed!* California Academic Press.

    Retrieved September, 1999 from: http://www.calpress.com/meta.html.

Facione, N., & Facione, P. (1996a). Assessment design issues for evaluating critical

    thinking in nursing. *Holistic Nursing Practice, 10*(3), 41-53.

Facione, N., & Facione, P. (1996b). Externalizing the critical thinking in knowledge

    development and clinical judgment. *Nursing Outlook, 44*(3), 129-136.

Facione, N., Facione, P., & Sanchez, C. (1994). Critical thinking disposition as a measure

    of competent clinical judgment: The development of the California critical

    thinking disposition inventory. *Journal of Nursing Education, 33*(8), 345-350.

Fearing, A. (1997). *Literature update on academic performance of minority*

    baccalaureate nursing students (ED3999920). Illinois.

Frierson, H., Malone, B., & Shelton, P. (1993). Enhancing NCLEX-RN performance: Assessing a three-pronged intervention approach. *Journal of Nursing Education, 32*(5), 222-224.

Grossman, D., Massey, P., Blais, K., Geiger, E., Lowe, J., Pereira, O., Stewart, A., Taylor, R., Filer, V., Nembhard, J., & Tally-Ross, N. (1998). Cultural diversity in Florida nursing programs: A survey of deans and directors. *Journal of Nursing Education, 37*(1), 22-26.

Hanks, C., & Lauchner, K. (1999). Indicia: Letters to the editor. *Computers in Nursing, 17*(6), 241-246.

Heupel, C. (1994). A model for intervention and predicting success on the National Council Licensure Examination for Registered Nurses. *Journal of Professional Nursing, 10*(1), 57-60.

Hewitt, P. (2002, October 30). Keeping score academically: National Merit semifinalists rate as cream of crop. *Houston Chronicle*, pp. 22A.

Kane, M. (1992). The assessment of professional competence. *Evaluation & the Health Professions, 15*, 163-182.

Kane, M. (1994). Validating interpretive arguments for licensure and certification examinations. *Evaluation & the Health Professions, 17*(2), 133-160.

Kane, M. (1997). Model-based practice analysis and test specifications. *Applied Measurement in Education, 10*(1), 5-18.

Kane, M., Crooks, T., & Cohen, A. (1997, March). *Justifying the passing scores for licensure and certification tests.* Paper presented at the AERA NCME Annual Meeting, Chicago, IL.

Kane, M., Kingsbury, C., Colton, D., & Estes, C. (1986). *A study of nursing practice and role delineation and job analysis of entry-level performance of registered nurses.* Chicago: National Council of State Boards of Nursing,

Kataoka-Yahiro, M., & Saylor, C. (1994). A critical thinking model for nursing judgment. *Journal of Nursing Education, 33*(8), 351-356.

Lauchner, K., Newman, M., & Britt, R. (1999). Predicting licensure success with a computerized comprehensive nursing exam: The HESI Exit Exam. *Computers in Nursing, 17*(3), 120-125.

Lockie, N., & Burke, L. (1999). Partnership in learning for utmost success (PLUS): Evaluation of a retention program for at-risk nursing students. *Journal of Nursing Education, 38*(4), 188-192.

Marshall, J. (1989). Student attrition: Is lack of support a key? *Nursing Outlook, 37*(4), 176-178.

Memmer, M. (1991). Retention of English-as-a-second-language (ESL) students: Approaches used by California's 21 generic baccalaureate nursing programs. *Journal of Nursing Education, 30*(9), 389-396.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (Third ed.). New York: American Council on Education.

Messick, S. (1994). The interplay of evidence and consequences in the validation of

 performance assessments. *Educational Researcher, 23*(2), 13-23.

Messmer, P., Abelleira, A., & Erb, P. (1995). Code 50: an orientation matrix to track

 orientation cost. *Journal of Nursing Staff Development, 11*(5), 261-264.

Miller, M., & Malcolm, N. (1990). Critical thinking in the nursing curriculum. *Nursing*

 and Health Care, 11(2), 67-73.

Morrison, S., Adamson, C., Nibert, A., & Hsia, S. (2002). HESI Exams: An overview of

 reliability and validity. Manuscript submitted for publication.

Morrison, S., Free, K., & Newman, M. (2002). Do progression and remediation policies

 improve NCLEX-RN pass rates? *Nurse Educator, 27*(2), 94-96.

Morrison, S., & Free, K. W. (2001). Writing multiple-choice test items that promote and

 measure critical thinking. *Journal of Nursing Education, 40*(1), 17-24.

Morrison, S., Smith, P., & Britt, R. (1996). *Critical Thinking and Test Item Writing*.

 Houston, TX: Health Education Systems, Inc.

National Council of State Boards of Nursing. (1999). *National Council detailed test plan*

 for the NCLEX-PN examination. Chicago, IL: National Council of State Boards of

 Nursing.

National Council of State Boards of Nursing. (2001a). *National Council detailed test plan*

 for the NCLEX-RN examination. Chicago, IL: National Council of State Boards of

 Nursing.

National Council of State Boards of Nursing. (2001b). *NCLEX-RN and NCLEX-PN: Examination statistics.* Retrieved July, 2001 from NCNSB Web site: http://www.ncsbn.org/research/nclexstats.nclex.asp.

National Council of State Boards of Nursing. (2002). *National Council of State Boards of Nursing (NCSBN) Web site.* Retrieved February, 2002 from: www.ncsbn.org.

National League for Nursing Accrediting Commission. (1999a). *Accreditation Manual 1999 for post secondary and higher degree programs in nursing.* New York: National League for Nursing Accrediting Commission.

National League for Nursing Accrediting Commission. (1999b). *Accrediting standards and criteria for academic quality of postsecondary and higher degree programs in nursing.* Retrieved July, 2000 from NLNAC Web site: http://www.accrediting-comm-nlnac.org/2am_stds&crit_fnl.htm#III.

Newman, M., Britt, R., & Lauchner, K. (2000). Predictive accuracy of the HESI Exit Exam: A follow-up study. *Computers in Nursing, 18*(3), 132-136.

Nibert, A., & Young, A. (2001). Predicting NCLEX success with the HESI Exit Exam: A third study. *Computers in Nursing, 19*(4), 172-178.

Paul, R. (1987). Dialogical thinking: Critical thought essential to the acquisition of rational knowledge and passions. In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: theory and practice* (pp. 127-148). New York: W. H. Freeman and Co.

Paul, R. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world.* Rohnert Park, CA: Center for Critical Thinking and Moral Critique.

Paul, R. (1994). *Critical thinking 2000: From an ideal evolves an imperative.* Paper presented at the Cutting Edge Issues: Critical Thinking Conference, University of Texas Medical Branch at Galveston.

Paul, R. (1999a). *A brief history of the idea of critical thinking.* The Center for Critical Thinking. Retrieved October, 1999 from The Center for Critical Thinking Web site: http://www.criticalthinking.org/University/univlibrary/crhistory.nclk.

Paul, R. (1999b). *Content is thinking: Thinking is content.* The Center for Critical Thinking. Retrieved October, 1999 from The Center for Critical Thinking Web site: http://www.criticalthinking.org/University/univlibrary/content.nclk.

Paul, R., & Elder, L. (1999). *Socratic Teaching.* The Center for Critical Thinking. Retrieved October, 1999 from The Center for Critical Thinking Web site: www.criticalthinking.org/CThink/k12/k12library/socratict.nclk.

Poorman, S., & Martin, E. (1991). The role of nonacademic variables in passing the National Council Licensure Examination. *Journal of Professional Nursing, 7,* 25-32.

Riner, M., Muller, C., Ihrke, B., Smoten, R., Wilson, M., Richardson, V., Stone, C., & Zwirn, E. (1997). Computerized NCLEX-RN and NCLEX-PN preparation programs: comparative review, 1997. *Computers in Nursing, 15*(5), 255-267.

Rodgers, S. (1990). Retention of minority nursing students on predominantly white

    campuses. *Nurse Educator, 15*(5), 36-39.

Rosella, J. (1994). The need for multicultural diversity among health professionals.

    Nursing and Health Care, 15(5), 242-246.

Ross, B., Nice, A., May, F., & Billings, D. (1996). Assisting students at risk: Using

    computer NCLEX-RN review software. *Nurse Educator, 21*(2), 39-43.

Rudy, E., Lucke, J., Whitman, G., & Davidson, L. (2001). Bechmarking patient

    outcomes. *Journal of Nursing Scholarship, 33*(2), 185-189.

Sax, G. (1997a). *Principles of educational and psychological measurement and*

    evaluation (Fourth ed.). Belmont, CA: Wadsworth Publishing Company.

Sax, G. (1997b). The validity of inferences obtained from tests and measurements. In G.

    Sax (Ed.), *Principles of educational and psychological measurement and*

    evaluation (Fourth ed., pp. 303-329). Belmont, CA: Wadsworth Publishing

    Company.

Schmidt, A. (2000). An approximation of an hierarchical logistic regression model used

    to establish the predictive validity of scores on a nursing licensure exam.

    Educational and Psychological Measurement, 60(3), 463-478.

Schwirian, P., Baer, C., Basta, S., & Larabee, J. (1978). *Prediction of successful nursing*

    performance (Publication No. 0-245-048). Washington, DC: US Government

    Printing Office.

Smith, P., & Crawford, L. (2002). The link between entry-level RN practice and the

    NCLEX-RN examination. *Nurse Educator, 27*(3), 109-112.

Symes, L., Tart, K., Travis, L., & Toombs, M. (2002). Developing and retaining expert

    learners: The student success program. *Nurse Educator, 27*(5), 227-231.

Taylor, C., Nahm, H., Loy, L., Harms, M., Berthod, J., & Wolfer, J. (1966). *Selection and*

    recruitment of nurses and nursing students. Salt Lake City: University of Utah

    Press.

Taylor, C., Nahm, H., Quinn, M., Harms, M., Mulaik, J., & Mulaik, S. (1965). *Report of*

    measurement and predication of nursing performance, Part I. Salt Lake City:

    University of Utah.

Tucker, R. W. (1999, September, 1999). *InterEd Web site*. InterEd, Inc., Phoenix, AZ.

    Retrieved September, 1999 from: http://www.intered.com.

United States Census Bureau. (2000). *Statistical Abstract of the United States. Number*

    196 Health professions schools - number, enrollment, and graduates: 1980 to

    1997. Retrieved July, 2000 from the United States Census Bureau Web site:

    http://www.census.gov/prod/www/statistical-abstract-us.html.

Vance, A. (1997). Strategies to assist students to be successful the next time around on

    the NCLEX-RN. *Journal of Nursing Education, 36*(4), 190-192.

Wall, B., Miller, D., & Widerquist, J. (1993). Predictors of success on the newest

    NCLEX-RN[1]. *Western Journal of Nursing Research, 15*(5), 628-643.

Waterhouse, J., Bucher, L., & Beeman, P. (1994). Predicting NCLEX-RN performance:

    cross-validating an identified classification procedure. *Journal of Professional*

    Nursing, 10(4), 255-260.

Waterhouse, J., Carroll, M., & Beeman, P. (1993). National Council Licensure

Examination success: Accurate prediction of student performance on the post

1988 examination. *Journal of Professional Nursing, 9*(5), 278-283.

Wilson, M. (1999). Using benchmarking practices for the learning resource center. *Nurse

Educator, 24*(4), 16-20.

Wolahan, C., & Wieczorek, R. (1991). Enrichment education: Key to NCLEX Success.

Nursing and Health Care, 12(5), 234-239.

Yearwood, E., Singleton, J., Feldman, H., & Colombraro, G. (2001). A case study in

implementing CQI in a nursing education program. *Journal of Professional*

Nursing, 17(6), 297-304.

APPENDIX A

Agency Permission

HESI

Anne Young, EdD, RN
Professor and Doctoral Program Coordinator
College of Nursing
Texas Woman's University
1130 John Freeman Blvd.
Houston, TX 77030-2597

Re: Approval for Dissertation Data Collection for Ainslie Nibert,

Dear Dr. Young:

This letter indicates my unconditional approval for Ms. Ainslie Nibert, MSN, RN, doctoral candidate enrolled in your program in the College of Nursing at TWU, to collect data for her dissertation study stored within the Health Education Systems Inc. (HESI) computerized database. Ms. Nibert will be using test scores obtained from administrations of the HESI Exit Exam (E^2) during academic year 1999-2000 and responses from the HESI Annual Survey of nursing program administrators obtained in 2001 regarding the NCLEX outcomes of their graduates that have been entered in the HESI database. No identifying information, such as student names, will be required for the data analysis. I am pleased that Ms. Nibert has chosen to conduct this study of the HESI E^2 to meet her dissertation requirements, and I look forward to reading the final version of the dissertation. Please do not hesitate to contact me at 713-838-7787, or via e-mail, susanm@hesitest.com if you have any questions regarding this approval.

Sincerely,

Susan Morrison, PhD, RN
President

2656 South Loop West, Suite 690 ● Houston. Texas 77054 ● 1.800.950.2728 ● fax: 713.838.0079 ● www.hesitest.com

113