PREDICTIVE ACCURACY OF THE HESI EXIT EXAM ON NCLEX-RN PASS RATES AND EFFECTS OF PROGRESSION POLICIES ON NURSING STUDENT EXIT EXAM SCORES

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

CAROLYN COHEN LEWIS, B.S., M.N.

DENTON, TEXAS

DECEMBER 2005

ACKNOWLEDGMENTS

With gratitude, the student wishes to acknowledge several individuals for their support and assistance. The dissertation committee members, Dr. Robin Britt and Dr. Carolyn Adamson, asked important questions to stimulate my thinking and provided valuable feedback on many aspects of this study. Dr. Anne Young, the dissertation committee chair, deserves particular recognition for her meticulous reading of my manuscript and kind recommendations. Thank you Dr. Susan Morrison for encouraging me to move to Houston, Texas, enroll in the Texas Woman's University doctoral program, for inspiring me with the idea of this research project and for granting me access to the HESI database. I wish to thank Dr. Ainslie Nibert for her guidance, support and encouragement in all the steps of the data collection process, for setting the example for my completion of the fifth HESI validity study, by publishing years III and IV; and for encouraging me to present my findings at various professional meetings. Dr. Mary Watson, thank you for challenging my data analysis techniques and for serving as a kind and nurturing mentor.

I wish to acknowledge my friends and colleagues at St. Luke's Episcopal Hospital (SLEH), especially Cheryl Lindy, Lavern Harris, Florence Schaefer, June Langford, Claudia Smith, and Peg Reiter for their encouragement; and all of the Texas Woman's University NURS 4806 Transitions senior students who completed clinical rotations at SLEH during the Fall 2002 through Fall 2005 semesters, for openly sharing their concerns, anxieties, and fears concerning TWU's Exit Exam graduation requirement. am grateful to my extended family including my mother, Louise Bennett Watkins, my

sisters Elizabeth Alice Killian, Constance Annette Adams, Jacqueline Edith Anderson, and Emily Ruth Blades, my brother Jared Benar Cohen, my in-laws Richard and Janice Cracroft; Jeff Cracroft, John and Louise Lewis, Katrina and Dean Peterson, Annette Brandon, and DaraLynn Lewis for their constant encouragement and support. A very special thank you is given to all the members of the Friendswood II Ward in the Friendswood, Texas Stake of the Church of Jesus Christ of Latter-day Saints for all their friendship, support and hours of self-less service to my family, during the past three and 1/2 years.

Finally, with eternal and heart felt love, I offer a tender thank you to my husband, Thomas Dorel Lewis, for his support and encouragement. I dedicate this dissertation to him and my three sweet daughters, Jennifer Diana, Sarah Louise, and Hannah Joy whose endless prayers, unending patience, and belief in me gave me the courage to keep going. Thank you for helping with all the extra chores that I normally would have completed; for putting my need to work on school assignments and dissertation above your own needs, and for always having faith in my ability to succeed in reaching this very challenging goal. I am truly grateful to you, my dear family, and will try in every way to make up for all the lost "togetherness" time we've endured because of the doctoral program. To all of the above mentioned, I am truly grateful to you, and will never forget all that you have given.

ABSTRACT

CAROLYN COHEN LEWIS

PREDICTIVE ACCURACY OF THE HESI EXIT EXAM ON NCLEX-RN

PASS RATES AND EFFECT OF PROGRESSION POLICIES

ON NURSING STUDENT EXIT EXAM SCORES

DECEMBER 2005

A descriptive retrospective research study was used to assess the predictive accuracy of the 2001-2002 Health Education Systems Incorporated (HESI) Exit Exam (E^2) on licensure exams, student NCLEX performance associated with various HESI scoring intervals, and the effects of benchmarking and remediation policies on student performance. Also, by means of a *t*-test, student scores from schools using designated E^2 benchmark scores for progression were compared with student scores from schools without such policies. *ANOVA* was used to evaluate differences in student performance based on ranges of a designated benchmark E^2 score as a 1) requirement for graduation 2) capstone course final grade weight; and 3) requirement for remediation; and to examine differences based on program establishment of a single benchmark consequence or combination of two or three benchmarking consequences.

The database sample consisted of 182 nursing schools with 9,695 students. E^2 was 97.8% accurate in predicting NCLEX success. Significant differences ($p <$ .0005) between scoring intervals indicated that NCLEX-RN failures increased as the scoring interval decreased. A significant difference existed between E^2 scores of students enrolled in nursing programs with benchmark policies and those enrolled in

programs without policies indicating that students perform better when consequences are attached than when there are no consequences ($p < .0005$).

When the E^2 score was required for graduation, schools with a benchmark set at 80 and below had higher E^2 scores than both the 85 ($p < .0005$) and 90 benchmark ($p = .001$). When E^2 scores were calculated as part of the course grade, there were no differences between the three groups (1-5%, 6-10%, and 16-20%). When the E^2 was required for remediation, schools with a benchmark of 90 had greater performance than the 85 ($p = .001$) or the 80 and below ($p < .0005$) benchmarks. Evaluation of student E^2 performance based on number of consequences found that students having two ($p < .0005$) or three ($p = .001$) consequences scored higher on the E^2 than students having only one consequence. These findings provide information faculties need to make decisions regarding use of E^2 scores as a benchmark in progression policies.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

A common problem shared by nursing faculties involves establishing program-specific policies related to testing and remediation designed to maintain desired National Council of State Boards of Nursing Examination for Registered Nurses (NCLEX-RN) pass rates for program graduates. State Boards of Nursing and the National League for Nursing Accreditation Commission (NLNAC) use achievement of acceptable NCLEX-RN pass rates as a major criterion measure of program effectiveness (Yocum, 1998). Poor NCLEX-RN pass rates negatively affect a school's reputation, ultimately placing the nursing program in jeopardy due to lack of interested applicants, falling enrollment, and loss of essential accreditation credentials and critical funding sources (Nibert & Young, 2001). Accurately predicting student success on the NCLEX-RN is highly desirable to nursing faculties and program administrations (Newman, Britt, & Lauchner, 2000).

An extensive review of the literature in search of predictor variables for NCLEX success revealed several comprehensive nursing examinations and traditional indicators, such as grade point average (GPA), have been evaluated for their accuracy in predicting NCLEX-RN outcomes for students in their final semester of a nursing program (Beeson & Kissling, 2001; Carpenter & Bailey, 1999). Comprehensive nursing exams were found to be more consistent in predicting NCLEX-RN outcomes because they are designed to measure overall ability, are standardized, and because they are administered closer to the time of NCLEX-RN testing.

The predictive accuracy of three comprehensive nursing exams, the Mosby Assess Test, the National League for Nurses (NLN) Achievement test, and the Health Education System Incorporated (HESI) Exit Exam (E^2) with NCLEX-RN success has been researched and reported in peer-reviewed nursing journals. The E^2 was chosen as the focus of the current study because it is the only comprehensive nursing exit exam with consistently high predictive accuracies (97.41%-98.3%) published over four consecutive years (Lauchner, Newman, & Britt, 1999; Newman et al., 2000; Nibert & Young, 2001; Nibert, Young, & Adamson, 2002). The E^2 is also the only computerized exam that is offered using the NCLEX CAT format, and compared to the Mosby Assess Test and the NLN Achievement test, the E^2 had the highest correlation with NCLEX outcome since initiation of a computerized NCLEX testing format (Barkley, Rhodes, & Dufour, 1998; Beeson & Kissling, 2001; Briscoe & Anema, 1999; Daley, Kirkpatrick, Frazier, Chung, & Moser, 2003; Lauchner et al., 1999; Newman et al., 2000; Nibert & Young, 2001; Nibert et al., 2002). Building predictive validity is a continuing process of reevaluation over time and circumstance of testing. One purpose of this study was to build on the validity and reliability work of previous studies.

There has been a trend for some nursing programs to implement progression or benchmark policies, associating consequences with student nursing exit exam scores. Nibert, Young, and Britt (2003) reported that of 149 programs that administered the E^2 during the 1999 – 2000 academic year, 45 (30.25%) indicated attainment of a faculty-specified benchmark E^2 was required for graduation. Students failing to achieve benchmarks at these 45 schools, faced one, two or all three consequences, including:

denial of eligibility for graduation, assignment of an incomplete or failing grade in the capstone course, and/or withholding of approval for NCLEX-RN candidacy. Program administrators reported progression policies specifying achievement of a minimally-acceptable E^2 benchmark score as a requirement for graduation were helpful in identifying academically at-risk students in need of remediation prior to NCLEX-RN candidacy (Nibert et al., 2003).

Susan Morrison of HESI reported (Susan Morrison, personal communication, January 17, 2004) that with establishment of strong predictive accuracy over four consecutive years, there has also been an increase in adoption of the HESI exams by nursing programs across the United States. Administrators and faculties of nursing schools are very interested in this tool that can so accurately predict NCLEX success. HESI database records indicated that use of the HESI exams by nursing schools increased from 245 in August 2000 to 546 by August 2003, an increase of 123% in three years (Morrison, Adamson, Nibert, & Hsia, 2004). With the increased E^2 usage, school administrators are requesting guidance regarding levels at which to set benchmarks, consequences related to E^2 scores, and implementation of progression policies (Susan Morrison, personal communication, August 2002). Inquiry also focuses on whether or not progression policies make a difference in student performance on the exam.

Problem of Study

This study was designed to build on prior predictive validity studies to evaluate scores reflecting a new scoring method whereby test items are individually weighted based on difficulty. Specifically, the accuracy of the E^2's new scoring method in

predicting NCLEX success and student NCLEX performance associated with various scoring intervals was investigated.

A second major focus of this study was the effect of established progression policies on student's E^2 performance. The E^2 scores of students attending nursing programs that used designated E^2 benchmark scores for progression were be compared with the E^2 scores of students attending nursing programs that did not have specific progression policies established that required specific E^2 benchmark scores.

In addition, this study evaluated differences in student performance based on ranges of a designated benchmark E^2 score as a 1) requirement for graduation 2) capstone course final grade weight; and 3) requirement for remediation. Finally, the study examined differences based on program establishment of a single benchmark consequence or combination of two or three benchmarking consequences.

Rationale for the Study

Over 87,000 United States educated nursing candidates take the NCLEX in a given year (NCSBN, 2004). It is critical that graduates be successful on this licensure examination. Many nursing faculty have come to rely on the ability of the HESI E^2 to predict NCLEX outcomes (Nibert, 2003). The E^2 has been a useful tool in assisting faculty to identify students who are at risk of failing the NCLEX during the last semester of their nursing program and in allowing faculty time to implement remediation strategies to better prepare students for NCLEX success (Nibert et al., 2003).

The E^2 is also appreciated for its ability to assist faculty in maintaining nursing program NCLEX pass rates. Two national nursing program accrediting bodies recommend that NCLEX pass rates should be maintained at a level of at least 92%

(AACN, 2002; NLNAC, 2004). Pass rates generally lower than 92% negatively indicate decreased program effectiveness in preparing program graduates (Yocum, 1998).

Not only does NCLEX failure negatively impact the reputation, enrollment and accreditation status of nursing programs; failure proves financially and personally devastating for the unsuccessful NCLEX candidates (Billings et al., 1996; Griffiths, Papastrat, Czekanski, & Hagan, 2004; McQueen, Shelton, & Zimmerman, 2004; Vance & Davidhizar, 1997). Also valuable work force resources invested in hiring and orienting new graduate nurses, who then fail the NCLEX, prove costly for health care employers (Messmer, Abelleira, & Erb, 1995).

Continued establishment of the E^2's predictive accuracy is a priority. According to Cronbach (1988), validity is based upon the many variables influencing the scores yielded by the test. Because of the dynamic, evolving nature of these variables including: nursing program curricula, nursing student population, E^2's scoring mechanisms, and the actual NCLEX-RN, regular confirmation of the E^2's ability to predict NCLEX success is a paramount requirement for its continued use among nursing programs.

Nursing faculties are using E^2 scores to predict NCLEX outcomes, and are also basing progression and remediation policies on required E^2 benchmark scores. The predictive accuracy of the exam is the basis for establishing progression policies. Without predictive accuracy confirmation, the E^2 and the progression policies tied to its predictive score would be of little benefit to nursing programs' NCLEX pass rates or their remediation strategies. Faculty who are currently using the E^2 as a graduation requirement, as a portion of the student's capstone course final grade, and / or as a

guide to require remediation, must be reassured that program progression policies are founded on evidence of the E^2's established ability to accurately predict NCLEX success.

In addition, findings of this study addressing how current progression policies and the three types of consequences associated with these policies, (ie., graduation requirement, capstone course final grade, or required remediation), are impacting student performance on the E^2, may be useful for nursing faculties considering implementation of a new benchmark policy. Results may shed light on whether or not a policy is even necessary for the support of their program outcomes, what type of policy or combination of consequences related to E^2 score would be most appropriate for their student population, or what specific E^2 score would be most beneficial for the established benchmark requirement. Findings may also assist faculties wishing to incorporate E^2 performance into calculation of the capstone course final grade, in determining how much weight the E^2 score should count toward the grade's calculation.

Theoretical Framework

The conceptual framework used in this study has two components: first, the predictive accuracy of the new (established in 2001) individually weighted item scoring method, and second, the impact of progression policies on student performance. Two theories establish the framework's study design for the new individually weighted item scoring method: predictive validity and classical test theory; and Weiner's attributional theory of motivation and emotion is the basis for the impact of progression policies on student performance.

The establishment of the E^2's predictive accuracy of NCLEX success is based upon a statistically oriented theoretical foundation called predictive validity, which concerns using an instrument to estimate some criterion behavior that is external to the measuring instrument itself (Nunnally & Bernstein, 1994). Developing a test that will estimate or predict performance on the NCLEX illustrates a predictive validity problem. The test ultimately chosen to predict outcomes is useful only insofar as it estimates NCLEX performance, defined as receiving a passing grade. Based on the measurable criterion, the validity of prediction is determined based on correlating scores on the predictor (E^2) with the criterion outcome (NCLEX). The size of the correlation directly indicates the predictive validity (Nunnally & Bernstein, 1994).

Predictive validity may refer to functional relations between a predictor and a criterion event occurring before, during or after the predictor is applied (Nunnally & Bernstein, 1994). For whichever time frame selected, the predictor measure is merely correlated to the criterion measure. The logic and procedure of determining predictive validity only require that data be available for correlation, not necessarily demanding a specific timeframe in which they were obtained (Nunnally & Bernstein, 1994). The nature of the question or problem of study determines when the two sets of measurements are obtained. For this study, the relationship between the predictor (E^2) and the criterion event (NCELX) is after the predictor is applied.

Predictive validity typically decreases as time elapses between measurements of the predictor and criterion (Nunnally & Bernstein, 1994). Since the E^2 score is obtained first, anything that happens to influence the criterion after the predictor scores are

obtained may potentially reduce the predictive validity. Theoretically, the longer the interval, the more opportunities there are for such events to occur (Nunnally & Bernstein, 1994).

Statistically, predictive validity is determined by, and only by, the degree of correlation between predictor and criterion (Nunnally & Bernstein, 1994). If the correlation is high, no other standards are necessary. However, sound theory and common sense are useful in selecting predictor instruments. It is important to emphasize that predictive validity is tied to a specific criterion; therefore, no theory assumptions can substitute for lack of correlation between predictor and criterion (Nunnally & Bernstein, 1994). Similarly, if multiple tests are evaluated for their predictive qualities of NCLEX performance, the test with the highest correlation with the criterion will be the most valid predictor.

Nunnally and Bernstein (1994) proposed additional considerations which determine usefulness of an actual test in particular situations including: size of the available pool of individuals, the proportion of people selected, and the difficulty of the situation. For this study, the size of the available pool of individuals will consist of approximately 14,727 nursing students who took the E^2 in their last semester of nursing school during the 2001 – 2002 academic year. Administrators of the nursing programs in the entire sample will be solicited for NCLEX outcomes of their students, with an anticipated response rate of approximately 60%. The E^2 is a comprehensive exam requiring high level critical thinking abilities to appropriately respond to application-level items (Morrison, Smith, & Britt, 1996). The mean item difficulties of the E^2, which corresponds to the proportion of examinees who answer the item correctly on a 0.00 –

1.0 scale, ranges from 0.3 - 0.7. Based on the proposed considerations of Nunnally and Bernstein (1994), the E^2 has the potential for yielding strong predictive accuracy results.

Cronbach (1990) suggested the overall strategy for employing predictor tests is to select tests that are considered parallel. The E^2 and NCLEX may be considered parallel exams if there is a strong correlation between them, and if the predictive accuracy from the first to the second is relatively high. Four previous studies published over four consecutive years have shown the E^2 to be highly predictive (97.41%-98.3%) of NCLEX (Lauchner et al., 1999; Newman et al., 2000; Nibert & Young, 2001; Nibert et al., 2002). The strong correlation of the two exams to date may be based on common properties shared between the two parallel exams. Both are comprehensive, computerized, criterion-referenced tests containing nursing application-based, critical thinking questions and are administered at the completion of a nursing program. Both allow for comparison with external norm groups, are classified as standardized tests within the content domain of nursing, are revised continuously based on reliability and validity indices, and both have the goal of predicting entry-level performance of registered nurses (Julian, Wendt, Way, & Zara, 2001; Nibert, 2003; Wendt, 2003).

The concept of parallel exams comes from the classical test theory, the basis for the type of testing used currently to evaluate nursing student learning, knowledge and skill acquisition. This theory is based on the true score model developed by Spearman in 1904 (Crocker & Algina, 1986) who proposed logical and mathematical arguments that test scores were fallible measures of human traits, and thus the observed correlation between fallible test scores (observed scores) is lower than the correlation between their "true objective values". The classical true score model suggests that there is always an

error of measurement associated with an examinee's observed test score. An error of measurement is the discrepancy between an examinee's observed test score and his true score, where the true score is the average of the observed scores obtained over an infinite number of repeated testings with the same test (Crocker & Algina, 1986). Hypothetically, if an examinee were to be tested an infinite number of times, the average of all the error scores would be equal to zero. With the error score equal to zero, the true score is the remaining score. This theory utilizes the concepts of statistical correlation to measure the amount of error in a test or research tool (Crocker & Algina, 1986).

Four key assumptions of the classical test theory are derived from the concepts of true and error scores (Crocker & Algina, 1986). First, for any given population of examinees, the mean error score and the correlation between true and error scores is zero. Second, the correlation between error scores from examinees' observed scores of two separate tests is zero. Third, an inverse relationship exists between random error and the reliability of a measuring instrument. As error increases, reliability decreases. Fourth, an inverse relationship exists between systematic error and validity. As the degree of systematic error in a measuring instrument increases, validity of the measure is decreased.

The classical test theory proposes that when two tests meet the requirements for parallel tests, it is possible to establish a mathematical correlation between them. The correlation derived is the measure of the accuracy of prediction of performance on NCLEX based on E^2 score. For this study, the students' E^2 scores that were greater than

90% were correlated with their NCLEX outcome to determine the amount of error in the prediction of NCLEX success.

Attributional Theory of Motivation and Emotion

The attributional theory of motivation and emotion by Bernard Weiner (1986) was the theoretical basis for five of the seven research questions in this research study. Weiner's theory builds upon the concepts of several previous motivation theories including Freud's psychoanalytic theory (1920), Hull's drive theory (1943), Atkinson and Feather's achievement theory (1966), Lewin's field theory (1935), Rotter's social learning theory (1954) and Tolman's purposive behavior theory (1925). This section will provide an overview of Weiner's attributional theory of motivation and emotion as shown in figure 1.

The theory has seven main components including outcome, outcome dependent affect, causal antecendents, causal ascriptions, causal dimensions, psychological consequences and behavioral consequences. The theory intricately describes the process of how an individual, either participant or observer, intellectually and emotionally processes and behaviorally reacts toward an unexpected outcome. The theory describes three aspects of cause identification, how the cause ultimately impacts the individual both cognitively and emotionally, and finally how this impact, then drives the individual's ultimate behavioral reaction toward the initial outcome.

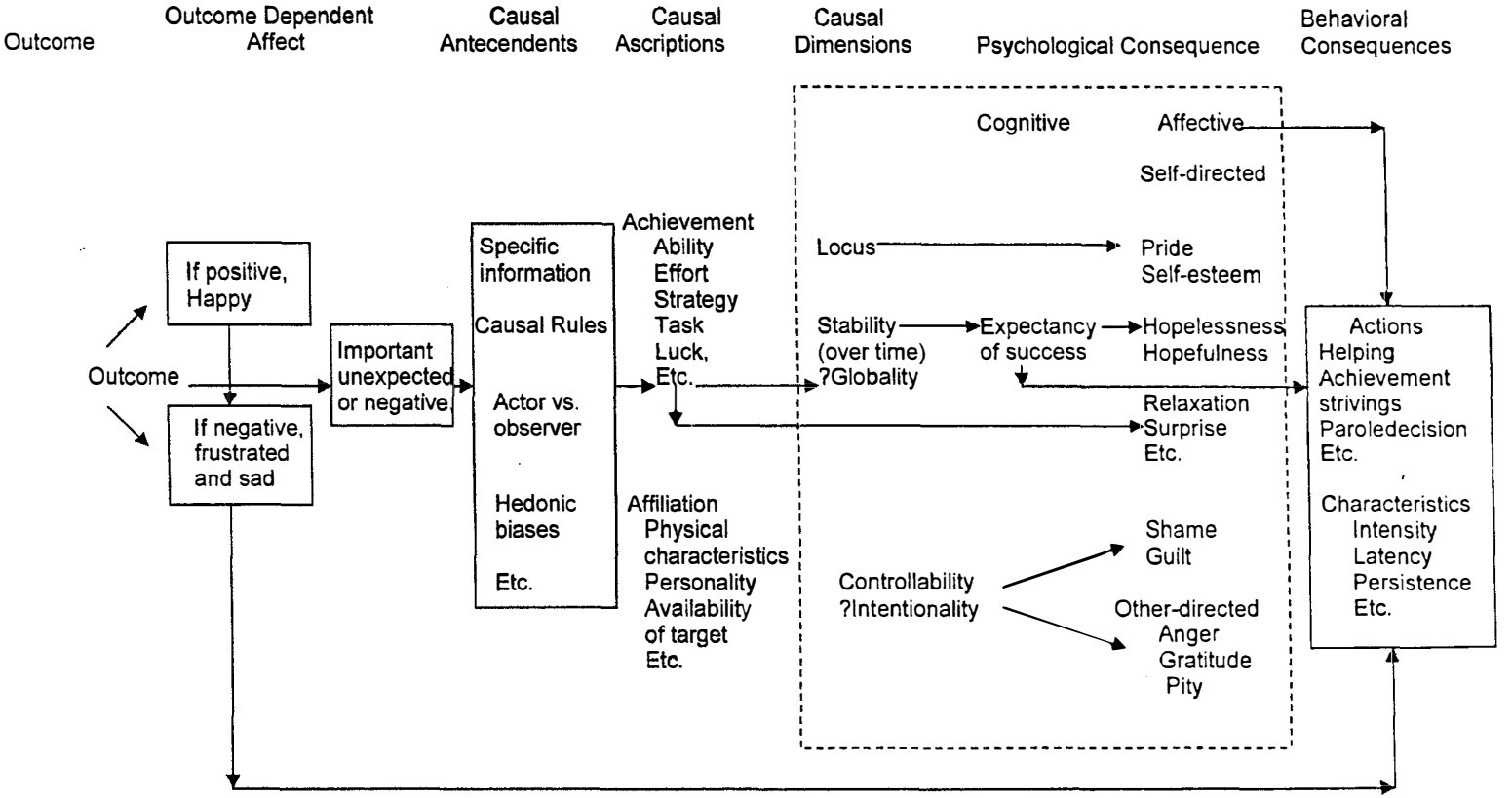

Figure 1. An attributional theory of motivation and emotion. *Note.* From An Attributional Theory of Motivation and Emotion (p. 16), by B. Weiner, 1986, New York City: Springer-Verlag New York. Copyright 1986 by Springer Science and Business Media. With kind permission. (See Appendix A).

Outcome and Outcome Dependent Affect

Weiner's (1986) attributional theory of motivation and emotion often begins with a completed event, an outcome. If that outcome is unexpected, important, and / or involves nonattainment of a desired goal, then an individual, either the participant or the observer displays an affect of surprise and either happiness, frustration, or sadness. It is apparent that the affect displayed is based on the unexpected outcome, and therefore, it is called the "*outcome dependent affect*."

Causal Antecedents

Immediately following the realization of the outcome, the individual initiates a causal search to answer "*why*" the outcome is the way it is. Many variables, called "*causal antecedents*," such as past history information, personal causal rules, and communications from others, influence how individuals arrive at their sought out conclusion. In other words, "*causal antecedents*" affect the results of the causal search, ultimately influencing what causal decision is reached. The causal decision reached is called the *causal ascription*.

Causal Ascriptions

Causal Ascriptions are chosen from many available causes to explain the outcome. They are the answers to questions such as "Why did I fail my exam?"; "Why is John doing so well on this job, after he performed so poorly at the previous ones?"; or "Why did Jill refuse my marriage proposal?" All of these "why" questions are imposed by the perceiver (either the individual or an observer) to account for the relation between an action and an outcome. Weiner et al. (1971) proposed that there are two motivational domains of causal ascriptions, the first has an achievement focus and the second has an

affiliation focus. Reasons for success or failure attributed to personal ability, effort, strategy, task difficulty, and luck are classified as achievement focused (achievement needs). Answers to "why" questions that include physical characteristics, personality, and availability of target fall into the affiliation focused (affiliation needs) motivational domain.

Cognitive Causal Dimension and Analysis

Each cause, once identified is judged (categorized) according to three dimensions: locus of causality, stability, and controllability. Locus of causality refers to whether the perceived cause is considered internal (a part / characteristic of the individual); or external (resulting from some outside force). When the locus of causality is perceived as internal, then emotions of pride and self-esteem result. If the cause is perceived as external, these emotions cannot occur (Weiner, 1986).

Stability refers to how constant a cause is in the impact of the outcome. cause to be considered stable, it must have constant capacity, such as ability. Unstable causal factors, such as effort or mood are perceived as more variable, changing from moment to moment or from one time interval to the next. Weiner's (1986) theory proposes that degree of stability of cause ultimately impacts perceived expectancy of success. If conditions surrounding the event (the presence or absence of causes) are expected to remain the same, then outcomes experienced on past occasions will be expected to recur. On the other hand, if causal conditions are perceived as likely to change, then the present outcome may not be expected to repeat itself in the future, or there may be uncertainty about subsequent outcomes.

Perceived stability of the cause leads to emotions of hopelessness or hopefulness. Emotional hopelessness occurs if the outcome is negative and the cause is perceived as stable and unchanging or if the outcome is positive, but the cause is perceived unstable and changing. On the other hand, hopefulness occurs if the outcome is positive and the cause is perceived as stable and unchanging or if the outcome is negative and the cause is perceived as unstable and changing (Weiner, 1986).

Controllability refers to feelings of individual control, or intentionality. If there is a belief that one is able to act upon the situation, then it is controllable. Uncontrollable refers to factors that are out of the control of the individual. The controllability dimension leads to emotions of shame, guilt, anger, gratitude, happiness or pity (Weiner, 1986). An example of how controllability leads to an emotional response is the student who perceives that the instructor who is grading their test is biased. The instructor's bias is perceived to be out of the control of the student, ie., a perceived uncontrollable factor in that student's test score. This lack of control gives rise to an angry emotional response within the student (Weiner, 1986).

The naturally occurring causal dimensional analysis of an unexpected outcome provides meaning or significance (cognitive consequences) via the three causal dimensions including locus, stability and controllability. Analysis also includes an affective component, the outcome's emotional reaction (affective consequences).

Affective Causal Dimension

Emotion is defined as a complex syndrome of many interacting factors having 1) positive or negative qualities of 2) a certain intensity, that 3) frequently are preceded by

an appraisal of a situation and 4) give rise to a variety of actions (Weiner, 1986). An important assumption guiding the approach to emotions is that how one thinks influences how one feels. Cognitions give rise to qualitative distinctions between feelings and therefore are responsible for the richness and diversity of emotional life. Cognitions typically precede and determine affective reactions, or in other words, perceptions of what caused a positive or a negative outcome in part determine the affective reactions to that outcome.

Emotion begins with the interpretation of an event as a success or a failure. This results in an overall positive (happy) or negative (sad, frustrated) affective reaction. If the outcome is negative, unexpected, or important, then attributional processes are elicited to determine the cause of that outcome. Causal attributions and their underlying properties of locus, stability, and controllability in turn generate differentiated affective reactions that are presumed to coexist with the initial broad emotional response. Each causal dimension is then linked with a dominant affective reaction (Weiner, 1986).

Behavioral Consequences

Weiner (1986) proposed that individual behavioral choices associated with an outcome occur based on individual differences in causal ascriptions and naturally occurring causal dimensional analysis. An example of how this works is when a student receives a failing grade for an important exam and asks "Why did I do so poorly?" The student clearly sees they studied and prepared a great deal, other classmates scored high and received passing scores. This particular student may conclude that the cause for poor performance on the exam is either attributed to lack of ability or lack of effort. Both ability and effort are internal locus, but they are different as to stability. Ability is an

example of a stable trait. Lack of effort is considered unstable. If a student attributes their poor performance on the exam to lack of ability, then, their perceived expectancy of success will be low for future exams. However, if a student attributes their poor performance on the exam to lack of effort, then their perceived expectancy of success will be high for future exams because they will feel that if their effort on future exams increases, so will performance. This example reinforces the concept that behavior intentions or consequences are due to expectations of future success; and high expectancies of success increase the intention to perform a particular behavior. Furthermore, the greater the perceived difficulty of the task required to achieve the desired outcome, the more effort may be perceived as necessary to attain the desired rewards. Conclusions drawn from causal dimensional analysis theoretically link causal ascriptions/attributions to three psychological aspects, thoughts, feelings and actions (Weiner, 1986).

Behavioral consequences involve taking action with varying degrees of intensity and persistence. When strong emotions are felt—and based on the perceived controllability, stability and locus of control of the causal attributions—the resulting behavior will show varying degrees of intensity. Causal dimensional analysis theoretically links attributions to cognitive and emotional psychological responses, which are then linked to achievement performance (Weiner, 1986).

The search for cause ultimately impacts the individual in two important ways—1) expectancy for success at a cognitive level; and 2) achievement strivings at an emotional level. Since attributions are linked to achievement performance, expectancy

of success and the resulting achievement strivings are the key determinants of motivation (Weiner, 1986).

Bidirectional Cause-Effect Linkages and Recurrent Feedback Loops

Figure 1, showing Weiner's theory moving in a consecutive pattern from an unexpected outcome toward ultimate behavioral consequences, is one representation of how the cognitive and emotional processes ultimately impact behavior. However, it is important to note that the cognitive and emotional processes do not necessarily have to occur in this manner. Weiner (1986) describes his attributional theory as being dynamically interactive, with reversible associations for at least some of the linkages in the theory. In reality, relationships between motivation and emotion have bidirectional cause-effect linkages. Attitudes affect behavior, and behavior influences attitudes; achievement or affiliation needs influence perception, but perceptions also generate achievement and affiliation needs; and desires can alter subjective expectancies, just as expectancy can modify desires. Relations between thinking, feeling and acting are not fixed within an orderly sequence (Weiner, 1986).

Motivation and the HESI Exit Exam

The E^2 is administered to students in their last semester of the nursing program as a way for faculty to assess student readiness for NCLEX-RN. Because the E^2 is used as a predictor of NCLEX-RN success, some nursing programs have implemented policies that use the E^2 score as a consequence or benchmark. Benchmark consequences include one or more of the following: 1) requiring a minimum E^2 score as a graduation requirement; 2) using the score to require a remediation course prior to

graduation; or 3) using the E^2 score as a portion of the capstone course grade (Nibert et al., 2003).

Since perceptions generate achievement and affiliation needs (Weiner, 1986); students who take the E^2 with a consequence attached may experience an increase in their achievement needs (causal ascriptions). Additionally, increased achievement needs may cause these students to perceive the E^2 as a high stakes exam. The perception of the high stakes exam, in turn, may cause these students to have greater achievement needs in relation to the exam compared to those students who may not perceive the E^2 as a high stakes exam because there are not consequences attached. Weiner's theory (1986) proposed that the greater the perceived difficulty of the task required to achieve the desired outcome, the more effort is perceived as necessary to achieve this outcome. According to Weiner's theory (1986), students who perceive this exam to be a high stakes exam may experience greater expectancy of success and achievement strivings than students who have no benchmark consequences tied to the exam.

Since all students take the E^2 in their last semester, they have experienced success in nursing school. Based on previous successes, (which represents stability during the nursing program) students feel positive and hopeful regarding their potential to perform well on the E^2. Their expectancy for success, tied also to their emotional hopefulness for success, influence their achievement strivings to succeed on the exam. However, it is proposed that the students who face the HESI E^2, perceiving it as a high stakes exam, may possibly make greater effort to perform well compared to students who take the exam without any consequences attached.

In summary, Weiner's attributional theory of motivation and emotion (1986) served as the theoretical basis for five of the seven hypothesis of this study proposing that a consequence or multiple consequences tied to the unexpected outcome on a high stakes exam, such as the E^2, stimulate(s) student cognitive and emotional reactions; resulting in student increased expectancy of success and enhanced performance strivings to succeed on the exam as manifested by higher exam scores.

Assumptions

Regarding determining the predictive accuracy E^2 based on classical test theory, the following assumptions were made in this study:

1. The E^2 and NCLEX are considered parallel exams based on a previously demonstrated strong correlation between them.
2. The E^2's ability to predict student performance on NCLEX is based on the size of the direct correlation between E^2 score and NCLEX outcome.

Three assumptions of Weiner's (1986) attributional theory of motivation and emotion were relevant to this study:

3. Achievement or affiliation needs influence perceptions.
4. Perceptions generate achievement needs.
5. Expectancy of success is associated with performance intensity of achievement strivings.

Research Questions

This study aimed to establish, the predictive accuracy of the E^2 and examine student NCLEX performance in relation to the various HESI scoring categories. Specifically, the research questions for this portion of the study were:

1. What is the predictive accuracy of the E^2 on NCLEX success using the weighted score calculation method for students in associate degree, baccalaureate degree and diploma registered nurse programs who took the test four to six months prior to graduation in the academic year 2001 – 2002.
2. Is there a difference in NCLEX outcomes for students whose E^2 scores are in the A/B, C, D, E/F, and G/H scoring intervals?

This study also looked at whether or not students perceive the HESI E^2 as a high stakes exam on the basis of having an existing benchmark or consequence associated with their final score; and whether or not this perception impacts their performance on the exam as reflected by their score. The research questions associated with this portion of the study included:

3. Are there significant differences between E^2 scores of students enrolled in baccalaureate (BSN), associate degree (ADN), or diploma programs that have adopted benchmark progression policies and those enrolled in these types of programs that have not adopted such policies?
4. Are there significant differences in E^2 scores among programs that specify one of three ranges of a designated benchmark E^2 score as requirement for graduation described as 90, 85, or 80 and below?
5. Are there significant differences in E^2 scores among programs that specify one of four ranges of capstone course final grade weights for the E^2 scores described as 1-5%, 6–10%, 11–15%, and 16–20%?

6. Are there significant differences in E^2 scores among programs that specify one of three ranges of a designated benchmark E^2 score for required remediation described as 90, 85, or 80 and below?
7. Are there significant differences in E^2 scores among students in nursing programs that use a single benchmarking consequence or a combination of two or three benchmarking consequences?

Definitions 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 the academic year 2001-2002, identified as Year V, which is the fifth year of study conducted on the E^2, within four to six months prior to graduation from an RN school of nursing.
2. Prediction scores were the total scores on the E^2 obtained from the HESI database. Scores 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 were calculated using the HPM. HESI scores were not expressed as percentages. Instead, these scores were calculated by applying the HPM to the raw score. The HPM considered 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 they 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. This study examined scoring intervals combined as follows: A and B (90-99.99); C (85-89.99); D (80-84.99); E and F (70-79.99); G and H ($\leq$ 69.99).

4. Predictive accuracy was described as the percentage of occurrences of the E^2 scoring interval successfully predicting a student's passage of NCLEX on the first attempt. Predictive accuracy for the students in the A/B HESI scoring category was determined by dividing the number that failed NCLEX by the number predicted to pass and subtracting from one.

5. Progression policy was a specific policy being implemented at a school of nursing at the time the E^2 is administered to senior students in their last semester of the nursing program, which used the HESI prediction E^2 score as the basis for a specific consequence or combination of consequences for the nursing student including: 1) determination of graduation eligibility or approval granted for NCLEX candidacy; 2) calculation of the capstone course final grade; and / or 3) determination of required participation in a specific remediation program. Specifics regarding individual program progression policies were obtained from school administrators when HESI staff requested NCLEX outcomes. Student NCLEX outcome and program progression policy information were collected prior to this study and stored in the HESI database.

6. E^2 benchmark scores were defined as the specific, numerical HESI prediction scores selected and reported by nursing program administrators as the minimally-acceptable scores students were required to attain before progression to graduation, approval for NCLEX candidacy, and / or participation in a specific remediation program was mandated, as specified by the program's specific progression policy.

7. Graduation eligibility refered to the student being allowed to graduate on the basis of their E^2 score having met or exceeded the benchmark requirement as mandated in their school's progression policy.

8. Approval for NCLEX candidacy referred to the letter the nursing program sends to the specific state board of nursing to give confirmation that the newly graduated nursing student is eligible to apply for registered nursing licensure and take NCLEX.

9. Capstone course final grade was the grade calculated for the final course of the student's nursing program.

10. Required remediation referred to the remediation strategies that the nursing student were obligated to participate in based on their E^2 score not meeting the designated benchmark according to the progression policy. Remediation strategies included elective courses, computer assisted instructional packages, NCLEX review books and other education materials used to assist students in improving weak content areas identified in the E^2 scoring reports.

Limitations

The following limitation must be considered when reviewing the findings of this study:

The research design lacked control over intervening factors (student remediation efforts, increased or decreased readiness for NCLEX, or increased or decreased anxiety, etc...) that may have occurred between the E^2 administration and administration of the NCLEX-RN. Such factors may have influenced the predictive accuracy of the E^2 by causing either increased or decreased student performance on the NCLEX-RN.

Summary

The fifth validity study of the HESI E^2 was designed to examine not only the accuracy of the E^2's new HPM scoring method based on individual item weighting, but also student licensure exam performance associated with the specific scoring intervals. A theoretical framework based on the statistical principle of predictive validity and classical test theory was used to establish the reliability and validity of the E^2 as a predictor of NCLEX outcomes.

An additional purpose of this study was to evaluate the effect of progression policies on nursing student E^2 scores. This portion of the study was theoretically based on the attributional theory of motivation and emotion which proposes that an increase in the perceived threat of failure can result in augmented performance. Students who take the E^2 knowing that their score will be used as the basis for a consequence as specified by their nursing program's progression policy, may perceive the E^2 as a high stakes exam and thus demonstrate increased performance as demonstrated by a higher means on the exam compared to students' E^2 scores from schools not enforcing any type of progression policy or benchmark consequence.

CHAPTER 2

REVIEW OF THE LITERATURE

Predicting success on the National Council Licensing Exam for Registered Nurses (NCLEX) has always concerned nurse educators in baccalaureate, associate, and diploma programs. Predictors are those factors that provide the best measure of a student's ability to succeed on the licensure examination. For many years nursing researchers have been conducting studies to identify the predictive variables indicative of a successful outcome on the nursing licensure examination.

A substantial number of studies have been conducted in attempt to identify significant variables that function as predictors for success on the exam. Prediction of NCLEX success has been and continues to be important for nursing programs since accreditation, enrollment, funding and reputation are based on NCLEX pass rates. Not only is indication of NCLEX success important for nursing programs—it is also important for the nursing graduates taking the exam. Graduates who fail are unable to obtain employment, and may experience a sense of low self-esteem (Billings et al., 1996). If predictors can be accurately identified, faculty may implement specific remediation plans prior to graduation addressing identified weaknesses of students and potentially improving NCLEX outcomes (Lauchner, Newman, & Britt, 1999b).

This literature review explores the quality, significance, and important implications of predictive variables for NCLEX success for three time periods related to the history of the licensure examination. Findings from four previous validity studies evaluating the predictive accuracy of the HESI E^2, which is the instrument under

investigation, are also addressed. This review of the literature offers background information about the variables of interest. It will portray and validate variables most predictive of NCLEX outcomes and conclude the need for additional study of the E^2 as one such predictor. In addition to predictor variables and validity, this review incorporates literature related to progression policies and benchmarking, and performance motivation will also be covered.

A comprehensive, computerized literature search was conducted for the year span 1995 – 2005 using the Cumulative Index to Nursing and Allied Health Literature (CINAHL), First Search, Journals @ Ovid Medline, WorldCat, Elton B. Stephens Company (EBSCO), Psych Info, and Educational Resources Information Center (ERIC). In addition, Dissertation Abstracts International was used to gather data from scholarly research, and citations were tracked from one study to another using the Web of Science database. General topics searched in various combinations included predictive accuracy, benchmarking, progression policy, nursing education, predictors of success on the NCLEX-RN, performance, Weiner, test performance, achievement, motivation, consequences, attributional theory, high stakes testing, higher education, and success. Every attempt was made to ensure a comprehensive review.

Literature Summarizing a Forty-Year Search for Predictor Variables

Because extensive research has been conducted that examines predictor variables or NCLEX success, several authors have published literature reviews summarizing the findings. Taylor et al. (1965) published a literature review measuring predictors of nursing performance conducted before 1965. Of 180 studies found, 77 were classified as prediction studies of student success in nursing programs, and among

these, 65 predictors of success were identified. The most common predictors were those related to standard indicators of scholastic aptitude or a nursing student's record of academic achievements while in the program.

Schwirian, Baer, Basta, and Larabee (1978) conducted a meta-analysis on literature published from 1965 to 1975 and reported seven studies focusing on predictors of students' success on the State Board Test Pool Examinations (SBTPE). Key predictors from this meta-analysis were the National League of Nursing (NLN) Achievement Test scores, theory course grades, and grade point averages (GPAs). The authors also found that although grades in students' clinical courses were not predictive of licensure exam performance, 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) extended the body of literature on NCLEX predictor variables by completing an analysis of research published from 1976 to 1998. The authors identified 67 research studies as well as additional NCLEX predictive factors from baccalaureate degree (BSN), associate degree (ADN) and diploma (PN) programs. Pearson correlation coefficients and varimax factor analyses were used to analyze predictive values of GPA's, Scholastic Aptitude Test (SAT) and American College Test (ACT) scores, and outcomes on standardized tests such as NLN examinations and the Mosby Assess Test. The authors found that: (1) all three types of programs shared similar predictors of NCLEX success; (2) academic factors and high school rank positively correlated with NCLEX success; (3) the NLN Baccalaureate Achievement Test and the Mosby Assess Test were the exams most frequently used to predict NCLEX

success, and (4) nursing theory courses in combination with NLN test scores appeared to be the best predictors of success. Despite these findings, Carpenter and Bailey (1999), still concluded that a consistently reliable predictor of licensure success had not been identified.

In 2004, Crow, Handley, Morrison, and Shelton surveyed 160 BSN programs in 38 states and the District of Columbia regarding selected admission, progression, and graduation variables that were used to predict NCLEX success during the 1999 academic year. Crow et al. (2004) found most programs used a comprehensive examination, (Mosby Assess test used most frequently); approximately one-third used cumulative GPA, (a mode of 2.5 as the GPA benchmark); and over one-third used specific course grades as a predictor of NCLEX success.

Table 1 provides a chronological summary of 42 studies, evaluating significant predictor variables of NCLEX success. Fourteen studies were identified by Carpenter and Bailey (1999) and 28 additional studies were released following Carpenter and Bailey's publication. The studies include data from 26 baccalaureate nursing (BSN), 14 associate degree nursing (ADN) and six practical nursing (PN) programs. Researchers studied 89 predictor variables and identified 40 (44.9%) that may be useful in accurately predicting NCLEX outcomes. The research presented was performed retrospectively and used samples of convenience ranging from 38 to 6,800 subjects.

Table 1 presents author(s) and year of publication, predictor variables studied, number of subjects, timeframe in which subjects took NCLEX, method(s) of statistical analysis, and significant predictor variables of NCLEX outcomes. Requisites for research to be included in this summary were (1) publication year during the period

Table 1

Chronological Summary of Predictor Variables of NCLEX Outcomes Examined in Research Literature 1990 – 2005

*(*Indicates research for the revised NCLEX blueprint and pass / fail scoring method which began in 1988).*

*(**Indicates research for the computerized NCLEX administration which began in 1994).*

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(1990) Dell & Valine	Collegiate GPA, SAT scores, ACT scores, self-esteem, & age.	78 ADN senior nursing students.	Multiple regression.	GPA = 58% of the variance in NCLEX-RN scores. All other variables were not significant.
(1990) Lengacher & Keller	Age, nursing program entrance & exit GPA's, ACT math & English; ACT composite, perception of role strain, nursing theory & clinical grades, NLN examinations.	146 ADN graduates; 1987 – 1988.	Pearson product moment correlation & step-wise multiple regression analysis.	Nursing program exit GPA (r = .71), ACT composite (r = .75), NURS 2712 (r = .77), NURS 2713 (r = .79), NLN Basics Two (r = .66), NLN Psychiatric (r = .70).
(1990) Schaal	Age, sex, race, transfer status, SAT math & verbal scores, entrance & exit nursing GPA; & GPA from biological, behavioral & physical sciences.	171 BSN graduates in 1983, 1985 & 1987.	Pearson product moment correlation & multiple regression analysis.	Age (r = .21), SAT verbal scores (r = .41), SAT math scores (r = .37). GPA at time of graduation (r = .57) & nursing core class GPA (r = .57) were the best predictors of NCLEX success.

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(1991) Foti & DeYoung	GPA overall, GPA nursing, GPA science, SAT verbal & math, NLN comprehensive exam, Mosby Assess Test Scores.	298 BSN graduates; 1985 - 1988.	Pearson product moment correlation & multiple regression analysis.	Mosby Assess Test (r = .66), GPA overall (r = .59), GPA nursing (r = .59), & NLN Achievement Test (r = .51). Most useful combination of variables: Mosby Assess Test, SAT verbal & overall GPA.
(1991) Horns, O'Sullivan, & Goodman	Age, sex, race, admission & exit GPA, grades for clinical & theory courses, & NLN comprehensive exam.	408 BSN graduates who completed NCLEX-RN 1985 – 1986.	Forward regression analysis.	Preadmission variables & sophomore course grades could be used to identify poor NCLEX-RN outcomes & plan early interventions.
(1991) Poorman & Martin	Quality point average (QPA), SAT, Test Anxiety Inventory (TAI),: total score & emotionality & worry subscales, self-perceived student grades, self-predicted NCLEX scores, concentration, negative & positive cognitions, physical symptoms, biggest worry related to NCLEX.	102 BSN senior nursing students. Date(s) not available, but actual NCLEX scores analyzed.	Pearson product moment correlation, multiple regression & chi-square analysis.	Test anxiety (r = -.31). Quality point average (r = .42), SAT (r = .30), Self perceived grades (r = .43), & self perceived NCLEX score (r = .32). Negative cognition was not inversely related to pass rate on the NCLEX-RN.
(1992) Fowles	ACT; GPA: prerequisite, science, liberal arts, 1st year & 2nd year; anatomy & physiology course grades; & Mosby AssessTest scores.	192 BSN graduates; 1985 - 1988.	Pearson product moment correlation & stepwise multiple regression analysis.	Mosby AssessTest (r = .79), Level 1 GPA (r = .74), & ACT composite (r = .53) & social science subscale (r = .55) scores.

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(1992) McClell&, Yang, & Glick	High School GPA, ACT score, chemistry, biological & social science GPA's, entrance & exit GPA, clinical & theory nursing course GPA, & Mosby AssessTest.	1,069 BSN graduates of nine Iowa schools from 1985 - 1988.	Pearson product moment correlation & step-wise multiple regression analysis.	ACT score (r = .48), entrance GPA (r = .41), Mosby AssessTest (r = .66), clinical GPA (r = .55), theory GPA (r = .53).
(1992) *Mills, Sampel, Pohlman, & Becker	Age, sex, high school GPA, ACT subscores, transfer status, nursing course grades & exit GPA.	531 BSN graduates. Took NCLEX 1982 – 1990.	Descriptive statistics & stepwise logistic regression.	Faculty identified students at risk for NCLEX failure 75% of the time. Age inversely related to NCLEX success. NCLEX-RN outcome best predicted at end of program.
(1992) *Mills, Becker, Sampel, & Pohlman	Age, sex, foreign educated, transfer GPA, & exit GPA.	328 BSN graduates; 1982 – 1990.	Stepwise logistical regression.	Exit GPA was a significant predictor of NCLEX outcome.
(1993) *Wall, Donald, & Widerquist	SAT score, high school rank, nursing entrance & exit GPA, science GPA, cumulative exit GPA, NLN course specific tests, & Mosby AssessTest.	92 BSN graduates; 1988 – 1991.	Discriminant function analysis, (chi-square) & t-test.	High school rank, science GPA, nursing entrance GPA, NLN achievement tests were significant predictors. Nursing exit GPA was most indicative of NCLEX success.

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(1993) *Waterhouse, Carroll, & Beeman	SAT scores, high school rank, course grades: physiology, pathophysiology & nursing; probation, transfer, changed major, & GPA: sophomore & exit.	257 BSN graduates; 1988 – 1990.	Pearson product-moment correlation, & discriminant function analysis.	Discriminant analysis correctly categorized 91.44% of subjects regarding NCLEX outcome. Exit GPA (r = .25), grades in senior level nursing course (r = .24), & SAT verbal (r = .23).
(1994a) Heupel	Nursing GPA: 1st, 2nd, 3rd & 4th years, exit GPA, five pre-requisite science courses & eleven selected nursing theory courses.	152 BSN graduates; 1985 – 1987.	Pearson product-moment correlation & multiple regression analysis.	Exit GPA (r = .60), junior level GPA (r = .51) & nursing theory course grades: N205 (r = .44), N311 (r = .60) & N421(r = .51).
(1994) *Waterhouse, Bucher, & Beeman	SAT verbal & math scores; high school rank, physiology grades & nursing course grades, GPA: sophomore & exit, & American Nursing Review (ANR) course.	135 BSN graduates; 1991 – 1992.	Descriptive statistics, Pearson product-moment correlation, t-test, chi-square & discriminant function analysis.	Discriminant analysis correctly categorized 84.00%. SAT verbal (t = 2.675); SAT math (t = 3.004); high school rank (t = 3.069); physiology grades (t = 3.243); & nursing theory grades (t = 7.780 & t = 3.139).
(1995) Drake & Michael	Grades in eight nursing theory & four biology-related courses; GPA: high school, community college, pre-requisite, cumulative nursing, cumulative biology & cumulative exit.	350 ADN graduates. Date(s) not available.	Multiple regression analysis.	Cumulative nursing GPA (r = .32).

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(1997) *Endres	GPA: entrance, medical-surgical, nursing core & cumulative; race, Mosby AssessTest; age; number of semesters; LVN status; "D" & "F" nursing course grades.	50 African-American, 50 foreign-born, & 50 white BSN graduates; 1987 – 1992.	Pearson product-moment correlation, Chi-square, two-way ANOVA, & discriminant analysis.	Mosby AssessTest (r = .49), entrance GPA (r = .24), & nursing GPA (r = .22). Mosby AssessTest percentile below 21 & D or F in a nursing course more likely to fail NCLEX. No differences in NCLEX pass rates between the three ethnic groups.
(1997) *Alexander & Brophy	High school rank, SAT scores, years of high school chemistry & math, admission status, GPA, number of credits, age, nursing & cognate course grades, Levels 1 & 2 GPA, & NLN Achievement Test.	188 ADN graduates; 1988 – 1994.	Chi-square, *t*-tests & logistical regression.	SAT verbal scores and nursing GPA best predictors. (no *r* provided). NLN Comprehensive Achievement Test scores in combination with three nursing courses predicted NCLEX success with 80.63% accuracy ($p < .0001$).
(1997) *Milan	Age, gender, race, ACT math & science, Pre-nursing GPA, GPA 1st yr, GPA 2nd yr., exit nursing GPA, NLN Med-surg test.	272 ADN graduates; 1992 – 1995.	Multivariate logistic regression.	Nursing exit GPA & NLN med-surg test. The multivariate equation developed using these two variables accurately predicted 86.40% of pass/fail performance on NCLEX-RN.

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(1998) *Arathuzik & Aber	Study Skills Self Efficacy Instrument (SSSE). Exit GPA, 1st language, family dem&s, emotional distress, & self confidence with critical thinking.	79 senior BSN students. Year(s) not available.	Descriptive statistics & Pearson product-moment correlation.	Exit GPA (r = .275); English as primary language (r = .253), self-confidence with critical thinking (r = .245), emotional distress (r = -.240) & family demands (r = -.293).
(1998) **Barkley, Rhodes, & Dufour	NCLEX-RN Risk Appraisal Instrument (RAI)t; nursing course grades; & NLN Adult Achievement test.	81 BSN graduates; 1997.	Descriptive statistics; Mann-Whitney U & Chi-square.	Pediatric nursing course (r = .59), Adult NLN Achievement Test (r = .59), the Psychiatric Mental Health nursing course (r = .58) & NCLEX-RN RAI score >1 (r = .78).
(1999) **Briscoe & Anema	Pre-admission GPA, clinical course failure, NLN test scores, age & race.	38 ADN graduates in 1997.	Pearson product-moment correlation.	NLN I exam (r = .48) , NLN II exam (r = .37), age (r = .37), & race (r = .47 for African descent).
(1999b) **Lauchner, Newman, & Britt	Health Education System Incorporated (HESI) Exit Exam (E^2) probability score.	2,809 RN & PN graduates; 1996 – 1997.	Chi-square test of significance.	99.49 % accurate in predicting NCLEX-RN outcomes when administered in monitored setting. Only 96.82% accurate in an unmonitored setting.

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(2000) **Lamm & McDaniel	PSB Aptitude for Practical Nursing Examination, race, overall GPA, course grades for anatomy & physiology courses, & high school graduation or GED diploma.	667 PN graduates; 1992 – 1996.	Point bi-serial correlation coefficients, non-parametric comparisons & logistical regression analysis.	Overall GPA (r = .43), race (r = .28), & the General Mental Ability (GMA) subscale of the PSB (r = .34).
(2000) **Newman, Britt, & Lauchner	HESI Exit Exam (E^2) probability score.	4,042 RN & PN graduates; 1997 – 1998.	Chi-square test of significance.	98.74 % (monitored setting) & 97.20% (unmonitored setting) accurate in predicting NCLEX-RN outcomes. Significantly more low-scoring students failed NCLEX than did high-scoring students.
(2000) **Nnedu	Age, race, gender, duration in nursing program, grades in pediatric & psychiatry nursing courses.	N not available BSN graduates. Year(s) not available.	Pearson product moment correlations, Chi-square, multiple regression & discriminant analysis.	Psychiatry nursing course grade most significant predictor. Older graduates have higher NCLEX pass rates than younger graduates. Whites pass NCLEX at higher rate (82%) than minority students (56%).

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(2001) **Nibert & Young	HESI Exit Exam (E^2) probability score.	6,560 RN & PN graduates; 1998 – 1999.	Chi-square goodness-of-fit.	97.78 % accurate in predicting NCLEX pass rates. Significantly more low-scoring students failed the NCLEX than did high-scoring students.
(2001) **Ostrye	Age, race, exit GPA, method of high school completion, Pell grant recipient, Aptitude for Practical Nursing Examination (PSB) scores, type of remedial basic skill courses taken, & nursing course grades.	226 PN graduates who failed NCLEX 1997 – 1999.	Chi-square, Wald statistic, forward inclusion logistic regression.	Exit GPA (β = 3.07) & PSB natural science subscale test (APNE) (β = .027).
(2001) **Gallagher, Bomba, & Crane	Course grades, Registered Nurse Entrance Exam (RNEE), Nursing Entrance Test (NET),	121 ADN students admitted in 1995.	Logistical regression analysis.	NURS 101 course grade predicted NCLEX success. RNEE predicted 50% probability of success in the NURS 101 course.
(2001) **Beeman & Waterhouse	Graduation year & age, sex, SAT scores, science & nursing course grades, number of C+ or lower in theory, number of B or lower in clinical, sophomore GPA, & 1st semester senior GPA.	236 BSN graduates; 1995 – 1998.	Pearson product-moment correlation & discriminant analysis.	Total number of C+ or lower grades earned in nursing theory (r = -.39) & grades in several nursing theory courses (r = .30 - .38). Greater than 93% of graduates were classified correctly.

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(2001) **Beeson & Kissling	Type of student (freshman admission, transfer, second degree); age; gender; pre-nursing, junior & senior nursing course grades; exit GPA & Mosby AssessTest scores.	505 BSN graduates; 1993 – 1998.	Mann-Whitney test, t-test, Chi-square, step-wise logistical regression & cross validation.	Combining the number of C grades or lower at the end of the junior year with the Mosby AssessTest score & age group predicted whether a student will fail NCLEX (66.7%).
(2001) **Percoco	Nursing course grades: pharmacology, psychology; preadmission grades: English & biology courses, number of remedial courses taken.	177 ADN graduates; 1991 – 1997	Logistic regression	Pharmacology grades predicted NCLEX success 78% of the time.
(2002) **Collins	Anatomy & Physiology I & II, chemistry, microbiology, entrance GPA, Combined nursing theory grades—Jr. yr.	159 ADN graduates; 1992 – 2001.	Pearson product-moment correlation & logistic regression.	Entrance GPA moderately significant in predicting NCLEX success. Three nursing theory grades—Jr. yr. were significantly predictive of NCLEX success, with the strongest overall predictor being pharmacology grade.
(2002) **Giddens	California Critical Thinking Disposition Inventory (CCDI), California Critical Thinking Skills Test (CCTST) & nursing exit GPA.	218 BSN students; 1998 – 2001.	Multiple regression analysis, *t*-test & discriminant analysis.	Prediction by the critical thinking measures CCTST & CCTDI were not any more significant than nursing exit GPA.

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(2002) **Henriques	Grade point average, Pre-nursing Guidance Test scores, Nelson-Denny Reading Test scores, California Critical Thinking Disposition Inventory (CCTDI) scores, Individual Learning Styles, California Critical Thinking Skills Test (CCTST) scores, & Diagnostic Readiness Test scores.	152 BSN students. Year(s) not available.	Multiple linear regression.	Nelson-Denny Reading Test & the Diagnostic Readiness Test.
(2002) **Nibert, Young, & Adamson	HESI Exit Exam (E^2) probability score.	6,800 RN & PN graduates; 1999 – 2000.	Chi-square goodness-of-fit.	98.46 % accurate in predicting NCLEX pass rates. More low scoring PN students failed NCLEX than low scoring RN students.
(2002) **Schafer	Age, HS GPA, ACT science & English subscores, Nurse Entrance Test (NET), Combined science prerequisite GPA, GPA 1st semester Jr, GPA 1st semester Sr.	327 BSN graduates. Year(s) not available.	Multiple & logistical regression.	Nursing GPA 1st semester Jr. best predictor. Several science courses improved odds of passing NCLEX. NET scores only slightly increased likelihood of NCLEX success.

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(2003) **Daley, Kirkpatrick, Frazier, Chung, & Moser	Comparison of age, gender, race, pre-requisite GPA, ACT scores, chemistry, anatomy, sociology, zoology, pathophysiology, senior medical-surgical nursing theory & clinical, cumulative GPA, Mosby AssessTest, HESI E^2 scores between successful & un-successful students.	121 BSN nursing students in 1999; & 103 BSN nursing students in 2000.	Independent *t*-test & Chi-square analysis.	Age, pre-requisite GPA, ACT, anatomy, pathophysiology, didactic & clinical medical-surgical grades & cumulative GPA were significantly different for those who passed & those who failed NCLEX. Mosby AssessTest efficiency 60%, & HESI efficiency 91%.
(2003) **Waterhouse & Beeman	Delaware Risk Appraisal Instrument (DRAI), nursing course grades & number of C+ grades or lower in nursing theory or B grades or lower in clinical.	538 BSN graduates; 1995 – 1998.	Pearson product-moment correlation & discriminant analysis.	Modified DRAI (r = -.315) much lower than the original RAI (r = -7.83) developed by Barkley et al., (1998). DRAI had almost same predictive validity as the second restorative nursing course (r = -.311). DRAI classified 74.1% of NCLEX success correctly in contrast to 81.6% who would be predicted by chance alone.

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
(2003) **Yin & Burger	Age, gender, race, type of student, LPN versus non-LPN, transfer status, & previous degrees; high school cumulative GPA; high school class rank; ACT composite score; college GPA & number of credit hours prior to entering nursing; Exit GPA; Grades in English, anatomy, physiology, chemistry & microbiology.	325 ADN graduates; 1997 – 2001.	Correlation analysis, *t*-test, Chi-square analysis, & logistic regression.	Entrance GPA, psychology, natural sciences average grade—significantly higher for group that passed NCLEX compared with group that failed (t = -2.30, p< .024). Entrance GPA correlated with NCLEX success (r = .15; p < .01). Entrance GPA combined with high school rank was significant in predicting NCLEX success.
**(2004) Crow, Handley, Morrison, & Shelton	Meta-analysis of all programs admission, progression, graduation requirements, prediction data, interventions & demographic data related to NCLEX pass/fail rates	160 BSN programs from 38 states & the District of Columbia.	Chi-square correlation analysis.	Standardized entrance examination (n = 12; p < ,0005); SAT scores (r = -0.4, p = .03, n = 34); NLN mental health exam scores (r = .55, p = .02, n = 18); NLN community health scores (r = .76, p = .02, n = 9); clinical proficiency scores (p = .03, n = 74); & exit examination scores (p = .05, n = 59).
**(2004) Haas, Nugent, & Rule	Gender, race, age, nursing exit GPA, admission GPA, cumulative GPA, verbal & quantitative SAT scores, merit scores & campus groups.	351 BSN Graduates; 1991-2001.	Discriminant analysis.	Nursing exit GPA (p < .0005). NCLEX "passers" had nursing exit GPA's 0.3 points higher than "failers." 61.2% of "true failers" & 71.0% of "true passers" identified ($\acute{a}$ =.10).

Author/Date	Predictor Variables	Sample	Statistical Methods	Significant NCLEX Predictors
**(2004) Seldomridge & DiBartolo	Native or transfer student status; entrance GPA; one semester of nursing GPA; exit GPA; grades earned in prerequisite & core nursing courses; test averages in beginning & advanced medical-surgical nursing courses; & performance on NLNCATBS.	186 BSN graduates; 1998 – 2002.	Logistic regression.	Combination of test average in advanced medical/surgical nursing & percentile score on the NLNCATBS predicted 93.3 % of NCLEX-RN passes & 50% of failures.

ACT = American College Testing exam
ADN = associate degree in nursing
ANOVA = analysis of variance
BSN = bachelor of science-nursing
E^2 = Exit Exam
GPA = grade point average
HESI = Health Education Systems, Incorporated
LPN = licensed practical nurse
NCLEX-RN = National Council Licensure Examination-Registered Nurse
NLN = National League for Nursing
NLNCATBS = National League for Nursing Comprehensive Achievement Test for Baccalaureate Students
QPI = Quality Product Index
SAT = Scholastic Aptitude Test
SBE = State Board Examination
SBTPE = State Board Test Pool Examination

of 1990 to 2005, (2) nursing students as subjects, and (3) research conducted to identify variables predictive of NCLEX outcomes.

A chronological 1990 to 2005 year span was chosen to evaluate differences in predictor variables during three separate time periods of NCLEX administration. From 1982–1987 the National Council of State Boards of Nursing (NCSBN) administered NCLEX used pencil and paper exams with individual scores. From 1988-1993, NCSBN retained the pencil and paper format, but changed the scoring to pass/fail. In 1994, NCSBN began administering NCLEX using computerized item-response testing with pass/fail scoring. Ten studies published from 1990 –1995 reported predictor variables for NCLEX in its initial pencil / paper, individual scoring format. Nine studies, dated from 1992 to 1998, are marked with an asterisk indicating evaluation of predictor variables of the second pencil / paper, pass / fail NCLEX format. Twenty-three studies marked with a double asterisk evaluated variables for the current, computerized item-response version of NCLEX, which began in 1994.

Identification of significant predictors of NCLEX outcomes cannot be achieved with only a mere chronological presentation of the nursing literature, a comprehensive categorization of variables and critical analysis of significant findings is required. An understanding of statistical methodology for predictive accuracy must also be demonstrated in order to understand the extent of significance offered by each significant predictor variable.

Predictive Accuracy

Predictive accuracy is a statistical correlation estimating the degree of preciseness with which a measure of one criterion will foretell the outcome of a second criterion thereby identifying the precise amount of uncertainty or lack-of-knowledge that exists between the two. Correlational analysis between variables specifies the form and degree of relationships among variables and constructs, and is a matter of degree expressed from 0 to 1 (or –1) (Nunnally & Bernstein, 1994). The size of the correlation is the predictive accuracy. The predictor variable with the highest and most consistent correlation to NCLEX outcomes is the variable that will most accurately predict NCLEX outcomes. Thus, an important aspect of this literature review is an analysis of how identified research studies determined correlation between predictor variables and NCLEX outcome, and a determination of which predictor variables were the most consistently significant.

The Pearson Product Moment Correlation was the basis for the statistical analysis in determining correlation for 17 of the 42 research studies. It is the most frequently used bivariate correlational procedure comparing only one predictor variable to the outcome criterion (NCLEX success) (Huck, 2000).

The Chi-Square (n = 16) provided basic correlational estimation for the nominal criterion of pass /fail NCLEX results on the basis of frequency. A t-test (n = 7) was used when the predictive criteria were interval data. Only the predictor variables found to be significantly correlated to NCLEX outcomes were further analyzed using multivariate statistical

methods to determine which of the significantly correlated variables, or combination of those variables most accurately predicted NCLEX outcomes.

The different types of multivariate analyses used in the literature were stepwise, forward or multiple regression (n = 14); stepwise, forward or logistical regression (n = 14), and discriminant function analysis (n = 9). Other statistical correlations included two-way ANOVA, Mann-Whitney U, Wald, and other non-parametric statistics, which were referred to, but not described. Most of the studies reported effect sizes as r^2 or R^2. None of the articles mentioned calculation of statistical power to estimate appropriate sample size. Alpha (p) values were deemed significant when found to be within the generally accepted statistical range of .001 - .05 in order to decrease the risk of type I error (Lipsey, 1990).

The Significance and Quality Predictor Continuum

Significance and quality of reported correlational and other significant findings were analyzed to determine which predictor variables, if any, were consistently closest to 1 on the –1 to 1 correlational continuum, or highest in their predictive accuracy. The closer to 1 that a variable achieves, or the higher the percentage of accuracy, the more likely it is that this variable will be a reliable and significant predictor of NCLEX success.

Table 2 places correlations (r) or other statistical findings significant predictor variables on a continuum of significance with their reported p-value, when available. Quality and stability of each predictor variable is presented via categorizing the significant variables by type (non-cognitive variables, nursing program pre-admission cognitive variables, nursing program progressive cognitive variables, and nursing program exit cognitive variables); and then reporting the frequency (n) that each variable

Table 2

Summary of Significance of Identified Predictors

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
Non-Cognitive Variables				
Age (n = 17)	Alexander & Brophy (1997)		*NOT PREDICTIVE*	
	Beeman & Waterhouse (1994)			*NOT PREDICTIVE*
	Beeson & Kissling (2001)			*NOT PREDICTIVE*
	Briscoe & Anema (1999) (p=.05)			r = .37, significant
	Dell & Valine (1990)	*NOT PREDICTIVE*		
	Endres (1997)		*NOT PREDICTIVE*	
	Haas et al. (2004)			*NOT PREDICTIVE*
	Horns et al. (1991)	*NOT PREDICTIVE*		
	Lengacher & Keller (1990)	*NOT PREDICTIVE*		
	Milan (1997)		*NOT PREDICTIVE*	
	Mills et al (1992) (p<.05)		Age inversely related to NCLEX outcomes	
	Mills et al (1992)		*NOT PREDICTIVE*	
	Nnedu (2000)			Older graduates have higher pass rates than do younger graduates.

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
	Ostrye (2001)			*NOT PREDICTIVE*
	Schaal (1990) (p<.003)	r = .21		
	Schafer (2002)			*NOT PREDICTIVE*
	Yin & Burger (2003)			*NOT PREDICTIVE*
Emotional distress / role strain (n = 2)	Arathuzik & Aber (1998) (p = .05)		r = -.240	
	Lengacher & Keller (1990)	*NOT PREDICTIVE*		
English as 1st language (n = 1)	Arathuzik & Aber (1998) (p = .05)		r = .253	
Family demands (n = 1)	Arathuzik & Aber (1998) (p = .05)		r = -.293	
High school rank (n = 4)	Alexander & Brophy (1997)		*NOT PREDICTIVE*	
	Wall et al. (1993) (p<.05)		Significant	
	Waterhouse et al. (1993)		*NOT PREDICTIVE*	
	Yin & Burger (2003)			Significant
Race (n = 9)	Briscoe & Anema(1999) (p=.05)			r = .47 for African descent
	Endres (1997)		*NOT PREDICTIVE*	
	Haas et al. (2004)			*NOT PREDICTIVE*
	Horns et al. (1991) (p<.01)	significant predictor		
	Lamm & McDaniel (2000) (p<.001)			r = .28
	Milan (1997)		*NOT PREDICTIVE*	

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
	Nnedu (2000)			Whites pass NCLEX at a higher rate (82%) than minority students (56%).
	Ostrye (2001)			*NOT*
	Schaal (1990)	*NOT PREDICTIVE*		
Self-confidence with critical thinking (*n* = 1)	Arathuzik & Aber (1998) (*p* = .05)		*r* = .245	
Self-perceived grades (*n* = 1)	Poorman & Martin (1991) (*p*<.05)	*r* = .43		
Self-perceived NCLEX scores (*n* = 1)	Poorman & Martin (1991) (*p*<.05)	*r* = .32		
Test anxiety inventory & subscales (*n* = 1)	Poorman & Martin (1991) (*p*<.05)	*r* = -.31		
Nursing Program Pre-Admission Cognitive Variables				
ACT composite scores (*n* = 5)	Dell & Valine (1990)	*NOT PREDICTIVE*		
	Fowles (1992) (*p*=.05)	*r* = .53		
	Lengacher & Keller (1990) *(p<.01)*	*r* = .75		
	McClelland et al.(1992) (*p*<.001)	*r* = .48		
	Yin & Burger (2003)			*NOT PREDICTIVE*

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
ACT social-science (*n* = 3)	Fowles (1992) (*p*=.05)	*r* = .55		
	Milan (1997)	*NOT PREDICTIVE*		
	Schafer (2002)			*NOT PREDICTIVE*
Diagnostic Readiness Test scores (*n* = 1)	Henriques (2002)			Primary predictor of NCLEX success.
Nelson-Denny Reading Test scores (*n* = 1)	Henriques (2002)			Primary predictor of NCLEX success.
Number of Cs or lower in pre-requisites (*n* = 1)	Seldomridge & DiBartolo (2004) (*p*=.002)			r = -.245
Pathophysiology grade(*n* = 2)	Waterhouse et al. (1993)		*NOT PREDICTIVE*	
	Seldomridge & DiBartolo (2004) (*p*<.0005)			r = .377
Pharmacology (*n* = 2)	Collins (2002) (*p*<.000)			Strongest predictor of NCLEX success.
	Percoco (2001)			Classified NCLEX success correctly 78% of the time.
Pre-nursing entrance GPA (*n* = 15)	Beeson & Kissling (2001)			*NOT PREDICTIVE*
	Collins (2002)			Moderately significant.
	Drake & Michael (1995)			*NOT PREDICTIVE*
	Endres (1997) (*p*<.05)		*r* = .24	
	Haas et al. (2004)			*NOT PREDICTIVE*

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
	Henriques (2002)			*NOT PREDICTIVE*
	Horns et al. (1991) ($p<.01$)	*NOT PREDICTIVE*		
	Lengacher & Keller (1990)	*NOT PREDICTIVE*		
	McClelland et al.(1992) ($p<.001$)	$r = .41$		
	Milan (1997)		*NOT PREDICTIVE*	
	Schaal (1990)	*NOT PREDICTIVE*		
	Seldomridge & DiBartolo (2004)			*NOT PREDICTIVE*
	Wall et al. (1993) ($p<.001$)		Significant	
	Waterhouse et al. (1993)		*NOT PREDICTIVE*	
	Yin & Burger (2003) ($p<.024$, $p<.01$))			($t = -2.30$, $r = .15$) For each .1 increase in GPA, the odds of passing NCLEX increased thrice.
PSB Aptitude for Practical Nursing ($n = 2$)	Lamm & McDaniel (2000) ($p<.001$)			General mental ability subscale of PSB $r = .34$,
	Ostrye (2001) ($p<.02$)			APNE $\beta = .027$
Registered Nurse Entrance Exam ($n = 2$)	Crow et al. (2004) ($p < .0005$)			Chi-square correlated to predict NCLEX outcomes.
	Gallagher et al. (2001) ($p < .05$)			Predicted 50% probability of success in NURS 101 course

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
SAT comprehensive scores (n = 6)	Beeman & Waterhouse (2001)			*NOT PREDICTIVE*
	Crow et al. (2004) ($p < .03$)			*r=-0.4*
	Dell & Valine (1990)	*NOT PREDICTIVE*		
	Poorman & Martin (1991) ($p<.05$)	$r = .30$		
	Wall et al. (1993)		*NOT PREDICTIVE*	
	Waterhouse et al. (1993)		*NOT PREDICTIVE*	
SAT verbal/math scores (n = 4)	Alexander & Brophy (1997)		*NOT PREDICTIVE*	
	Foti & DeYoung (1991) ($p<.001$)	Useful when combined with Mosby Assess Test		
	Haas et al. (2004) ($p<.001$)			Passers exceeded failers in verbal scores.
	Schaal (1990) ($p<.001$)	verbal $r = .41$ math $r = .37$		
Science, social / behavioral GPA (n = 2)	Fowles (1992) ($p=.05$)	$r = .55$		
	Schaal (1990)	*NOT PREDICTIVE*		
Science, cumulative GPA (n = 6)	Beeman & Waterhouse (2001)			*NOT PREDICTIVE*
	Drake & Michael (1995)			*NOT PREDICTIVE*
	Foti & DeYoung (1991)	*NOT PREDICTIVE*		

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
	Schafer (2002)			*NOT PREDICTIVE*
	Wall et al. (1993)		Significant	
	Yin & Burger (2003)			*NOT PREDICTIVE*
Nursing Program Progressive Cognitive Variables				
Fundamentals theory grades (n = 1)	Gallagher et al. (2001) ($p < .05$).			NURS 101 predicted NCLEX success.
Level 1 Nursing GPA (n = 7)	Alexander & Brophy (1997)		*NOT PREDICTIVE*	
	Collins (2002)			Strongest variables to predict NCLEX success.
	Fowles (1992) (p=.05)	r = .74		
	Heupel (1994)	*NOT PREDICTIVE*		
	Milan (1997)		*NOT PREDICTIVE*	
	Schafer (2002)			Best predictor of NCLEX success.
	Seldomridge & DiBartolo (2004)			*NOT PREDICTIVE*
Medical-surgical theory grades (n = 10)	Alexander & Brophy (1997)		*NOT PREDICTIVE*	

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
	Beeman & Waterhouse (2001)			several theory courses contributed to prediction, r = .30 - .38
	Drake & Michael (1995)			*NOT PREDICTIVE*
	Endres (1997)		*NOT PREDICTIVE*	
	Heupel (1994) (p<.0001)	Three nursing theory courses significant: r = .44, r = .60; & r = .51		
	Horns et al. (1991)	*NOT PREDICTIVE*		
	Lengacher & Keller (1990) *(p<.05)*	Two nursing theory courses had significant correlations: r = .77, r = .79		
	Mills et al (1992)		*NOT PREDICTIVE*	
	Nnedu (2000)			Psychiatry nursing course grade most significant predictor of NCLEX success
	Ostrye (2001)			*NOT PREDICTIVE*
Pharmacology grade (n = 1)	Collins (2002) (p<.000)			Strongest predictor of NCLEX success overall.

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
3rd year theory course grade (*n* = 4)	Barkley, Rhodes et al. (1998) (*p*=.001)			Pediatrics (*r* = .59) Psychiatrics (*r* = .58)
	Beeson & Kissling (2001)			*NOT PREDICTIVE*
	Heupel (1994) (*p*<.0001)	*r* = .51		
	Waterhouse et al. (1993)		*NOT PREDICTIVE*	
4th year, 1st semester theory grade (*n* = 4)	Beeman & Waterhouse (2001)			*NOT PREDICTIVE*
	Beeson & Kissling (2001)			*NOT PREDICTIVE*
	Heupel (1994)	*NOT PREDICTIVE*		
	Waterhouse et al. (1993) (*p*<.05)		*r* = .24	
Test average-Adult Health II course (*n* = 1)	Seldomridge & DiBartolo (2004) (*p*<.0005)			r = .307
Number of Cs or Ds & Fs earned in nursing courses (*n* = 4)	Barkley, Rhodes et al. (1998) (*p*=.001)			*NOT PREDICTIVE*
	Beeman & Waterhouse (2001)			Probability of NCLEX high with > 3 Cs in nursing theory courses (*r* = .39)
	Endres (1997) (*p*<.05)		D or F course grade more likely to fail NCLEX	
	Seldomridge & DiBartolo (2004) (*p*<.0005)			# of C's (r = -.34)

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
Nursing clinical course grade (*n* = 6)	Beeman & Waterhouse (2001)			*NOT PREDICTIVE*
	Crow et al. (2004) (*p* = .03)			Clinical proficiency correlated with NCLEX pass rates.
	Drake & Michael (1995)			*NOT PREDICTIVE*
	Horns et al. (1991) (*p*<.01)	*NOT PREDICTIVE*		
	Lengacher & Keller (1990) (*p*<.001)	*r* = .79		
	McClelland et al.(1992) (*p*<.001)	*r* = .55		
Nursing Program Exit Cognitive Variables				
Cumulative college exit GPA (*n* = 16)	Arathuzik & Aber (1998) (*p* = .05)		*r* = .275	
	Dell & Valine (1990) (*p*<.001)		GPA = 58% of variance	
	Endres (1997)		*NOT PREDICTIVE*	
	Foti & DeYoung (1991) (*p*<.0001)	*r* = .59		
	Haas et al. (2004)			*NOT PREDICTIVE*
	Heupel (1994) (*p*<.0001)	*r* = .60		
	Horns et al. (1991) (*p*<.01)	significant predictor		

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
	Lamm & McDaniel (2000) ($p<.001$)			$r = .43$
	Lengacher (1990) ($p=.01$)	$r = .71$		
	Milan (1997)		Combined with NLN med-surg predicted 86.40% NCLEX pass/fail.	
	Mills et al (1992) ($p<.05$)		75% of variance	
	Mills et al (1992a) ($p<.05$)		Significant predictor	
	Ostrye (2001) ($p<.02$)			$\beta = 3.07$
	Poorman & Martin (1991) ($p<.05$)	$r = .42$		
	Wall et al. (1993)		*NOT PREDICTIVE*	
	Waterhouse et al. (1993) ($p<.05$)		$r = .25$	
Cumulative nursing exit GPA ($n = 11$)	Alexander & Brophy (1997) ($p<.001$)		Significant predictor	
	Beeson & Kissling (2001)			*NOT PREDICTIVE*
	Drake & Michael (1995) ($p<.001$)		$r = .32$	
	Endres (1997) ($p<.05$)		$r = .22$	
	Foti & DeYoung (1991) ($p<.0001$)	$r = .59$		

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
	Giddens (2002)			Prediction by the critical thinking measures CCTST and CCTDI not any more significant than nursing exit GPA.
	Haas et al. (2004) ($p<.0005$)			Passers had 0.3 greater than failers.
	McClelland et al. (1992) ($p<.001$)	$r = .53$		
	Schaal (1990) ($p<.001$)	$r = .55$		
	Seldomridge & DiBartolo (2004)			*NOT*
	Wall et al. (1993) ($p<.001$)		Significant	
Health Education Systems Inc. (HESI) Exit Exam (E2) ($n = 5$)	Daley et al. (2003) ($p<.001$)			Combined pass/fail predictive efficiency 91%
	Lauchner et al. (1999) ($p=.05$)			99.49 % predictive accuracy monitored & 96.8% accuracy unmonitored
	Newman et al. (2000) ($p=.05$)			98.74% predictive accurate monitored & 97.20% accuracy unmonitored

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
	Nibert & Young (2001) (p=.05)			97.78% predictive accuracy of NCLEX success
	Nibert et al. (2002) (p=.05)			98.46% predictive accuracy of NCLEX success
Mosby AssessTest Scores (n = 7)	Beeson & Kissling (2001) *(p<.001)*			Significant when combined with number of C grades and age group—66.7% accurate
	Daley et al. (2003) (p<.001)			Combined pass/fail predictive efficiency 60%
	Endres (1997) (p<.05)		Percentile rankings< 21 more likely to fail NCLEX. (r = .49)	
	Foti & DeYoung (1991) (p<.0001)	r = .66		
	Fowles (1992) (p=.05)	r = .79		
	McClelland et al. (1992) (p<.001)	r = .66		
	Wall et al. (1993)		*NOT PREDICTIVE*	

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
NCLEX-RN Risk Appraisal Instrument (RAI) or Delaware RAI (*n* = 2)	Barkley, Rhodes et al. (1998) (*p*=.001)			*r* = .78 if score > 1
	Waterhouse & Beeman (2003) (p<.001)			*r* = -.315, but compared to -.261 of a single course grade, not significantly better. Classified success on NCLEX 74.1% compared to 81.6% by chance alone.
NLN Achievement Test (*n* = 9)	Alexander & Brophy (1997) (*p*<.0001)		80.63% variance combined w/ 3 nursing courses.	
	Barkley, Rhodes et al. (1998) (*p*=.001)			*r* = .59
	Briscoe & Anema (1999) (*p*=.01)			NLN I Exam *r* = .48, NLN II Exam *r* = .37
	Crow et al. (2004) (*p* = .02)			NLN Mental Health, *r* = .55; NLN Community Health, *r* = .76.

Significant Predictors (how many times measured)	Reference & Significant *p* value, if any	Significantly Predicted NCLEX Outcomes Before 1988	Significantly Predicted NCLEX Outcomes 1988 – 1993	Significantly Predicted NCLEX Outcomes 1994 – 2003
	Foti (1991) ($p<.0001$)	$r = .51$		
	Horns et al. (1991) ($p<.01$)	Significant predictor		
	Lengacher & Keller (1990) ($p=.01$)	NLN Basics II $r = .66$, NLN Psychiatric $r = .70$		
	Milan (1997)		Combined with nursing exit GPA, predicted 86.40% NCLEX pass/fail.	
	Wall et al. (1993) ($p<.001$)		Significant	
NLNCATBS ($n = 1$)	Seldomridge & DiBartolo (2004) ($p<.0005$)			r = .452; When combined with pathophysiology grades, predicted 93.3% of successes and 50% of failures.

was studied and found to be significant. Researchers' significant findings are placed in the table according to the three NCLEX timeframes previously described. This provides a visual representation of frequency and chronological significance of findings. Table 2 reveals the true significance of the predictor variables and is the basis of understanding the true state of the science in predictive accuracy of NCLEX success.

It is interesting to note the actual correlational values of the most commonly studied and most relied upon predictor variables, and to realize these have served as the basis of identification and intervention strategies for at-risk students since 1990 despite their weakness. This chronological depiction of predictor variables shows the duration of impact significant findings can have in nursing education, even when the actual correlational values are mediocre at best. The frequency with which each variable is analyzed for its predictive abilities compared to the actual findings, shows that many of the variables were found as often insignificant as they were significant. Thus frequency of the variable in the research is by far the most conclusive as to the inconsistency of the research studies. A more in-depth discussion follows for each category of predictor variables.

Non-cognitive variables. Fifty-four non-cognitive variables were evaluated in 25 research studies including admission status, age, biggest worry related to NCLEX, California Critical Thinking Disposition Inventory (CCDI), California Critical Thinking Skills Test (CCTST), campus groups, changed major, concentration, duration in nursing program, emotional distress / role strain, English as first language, family demands, foreign educated, gender, graduation year, high school graduation or GED diploma, high school rank, licensed vocational nurse status, negative and positive cognitions, number

of credits, number of remedial courses, number of semesters, Pell grant recipient, physical symptoms, pre-nursing guidance test scores, probation, race, self-confidence with critical thinking, self esteem, self-perceived NCLEX scores, Study Skills Self Efficacy Instrument, test anxiety inventory and subscales, transfer or not transfer, type of student (freshman admission, transfer, or second degree), and years of high school chemistry and math. Of those listed, only 10 (38%) were found statistically significantly correlated with NCLEX success including age, emotional distress / role strain, English as first language, family demands, high school rank, race, self confidence with critical thinking, self-perceived grades, self-perceived NCLEX scores, test anxiety inventory and subcales. Age (n = 17), race (n = 9), and high school rank (n = 4) were the most frequently analyzed non-cognitive variables.

Age, although studied 17 times, was only found to be significant in four (23.5%) investigations. Schaal (1990) found age to be positively correlated with NCLEX outcomes (r = .21, p <.003) during the first NCLEX timeframe; as did Briscoe and Anema (r = .37, p = .05) in 1999, during the third NCLEX timeframe. Nnedu (2002) reported older graduates to have higher pass rates than younger graduates. In contrast, Mills et al., (1992a) found age to be inversely correlated to NCLEX outcomes. The disagreement in findings, and low number of significant findings compared to times studied tends to nullify correlations.

Race as a predictor variable for computerized NCLEX testing was found to be significant (r = .21, p =.05) by Briscoe and Anema (1999) and Nnedu (2000) who reported that students with African descent had significantly higher correlations with NCLEX failure than their Caucasian counterparts. Lamm and McDaniel (2000) also

reported race as significantly correlated with NCLEX outcomes (r = .28, p <.001). Horns et al. (1991) studied race during the first NCLEX time period and found it a significant (p < .0001) predictor. Race was not found as a significant predictor of NCLEX outcomes by Endres (1997), Haas (2004), Milan (1997), Ostrye (2001), and Schaal (1990).

High school rank was not studied during the first NCLEX timeframe. It was studied three times during the time frame when NCLEX was changed to pass/fail by Alexander and Brophy (1997), Wall et al. (1993) and Waterhouse et al (1993); and found to be significant (p < .05) only by Waterhouse et al. Yin and Burger (2003), (p not provided) reported high school rank as a significant predictor of success on the computerized NCLEX when combined with entrance GPA.

Arathuzic and Aber (1998) found English as a 1st language (r = .253, p = .05); and self-confidence with critical thinking (r = .245, p = .05), significantly correlated to NCLEX success, and reported family demands (r =-.293, p = .05) and emotional distress / role strain (r = -.240, p = .05) as non-cognitive variables inversely related to NCLEX success. However, Lengacher (1990) did not find emotional distress / role strain as a significant predictor. Poorman and Martin (1991) found self-perceived grades (r = .32, p < .05) and self-perceived NCLEX scores (r = .32, p < .05) to correlate with NCLEX success and test anxiety (r = -.31, p < .05) to inversely correlate.

Overall, researchers who have explored non-cognitive variables for prediction of NCLEX outcomes in order to provide interventional development addressing identified student weakness have had very limited success. Lack of significance and reproducibility of findings may be attributed to the subjective nature of non-cognitive variables and inability to predict or control for confounding variables. Faculty

intervention addressing age, race, high school rank, self-confidence, family demands, or anxiety and emotional distress may also prove to be ineffective, or minimally helpful at best since these variables are not ones that can be controlled. Strong correlations and stable predictors have yet to be found among non-cognitive variables.

Cognitive pre-admission variables. Variables that were used as admission criteria into nursing programs were categorized as cognitive pre-admission variables. Fourteen of 31 measures determined by at least one researcher to be a significant predictor of NCLEX success included ACT composite, ACT social-science, Diagnostic Readiness Test, Nelson-Denny Reading Test, number of C's or lower in pre-requisites, pathophysiology course grade, pharmacology course grade, pre-nursing entrance GPA, PSB Aptitude for Practical Nursing, Registered Nurse Entrance Exam, SAT comprehensive, SAT verbal/math, combined science & social science GPA, and cumulative science GPA. Researchers also evaluated 17 additional pre-admission variables that were not found to be predictive including ACT math, ACT English, anatomy and physiology course grades, biological science GPA, chemistry course grade, English, high school GPA, liberal arts GPA, merit scores, microbiology, number of C's or lower in pre-requisites, Nursing Entrance Test, physical science GPA, psychology course grade, remedial reading, science courses and statistics course grade.

Although less than half of the pre-admission variables studied were found to be significant, the quality of the significance for the cognitive pre-admission variables, on the whole tended to be more strongly correlated (r = -.15 - .75) as compared to the non-cognitive variables. Three pre-admission variables researched frequently were ACT scores (n = 8), pre-nursing entrance GPA (n = 15) and SAT scores (n = 10). Fowles,

(1992) (r = .53 & r = .55, p = .05), Lengacher and Keller (1990) (r = .75, p < .01), McClelland, Yang and Glick (1992) (r = .48, p < .001) reported significance with ACT composite scores prior to the NCLEX pass / fail changes in 1988; but ACT composite or social-science scores were not found to be significant by Dell and Valine (1990), Milan (1997), Schafer (2002), or Yin and Burger (2003).

Pre-nursing entrance GPA was most strongly correlated (r =.41) with NCLEX success prior to 1988 when NCLEX changed to the pass / fail format (McClelland, Yang, & Glick, 1992). However, for this same time period, Horns et al. (1991) and Lengacher and Keller (1990) did not find significance. Endres (1997) (r = .24, p < .05) and Wall et al. (1993) (p< .001), found GPA to be a significant predictor variable, when the NCLEX format changed to pass / fail; but Milan (1997) and Waterhouse (1993) did not. Collins (2002), Drake and Michael (1995), Haas (2004), Henriques (2002), Seldomridge and DiBartolo (2004), and Yin and Burger (2003) evaluated pre-nursing GPA as a predictor variable for the computerized NCLEX, but only Yin and Burger (r = .15, p < .01) and Collins (p not provided) found it to be a small to moderately significant predictor.

During initial administration of NCLEX, when scores were provided, Foti (1991) identified SAT scores to be useful as a predictor for NCLEX success when combined with the Mosby Assess Test; Poorman and Martin (1991) (r = .30, p < .05), and Schaal (1990) (r = .41 & r = .37, p < .001) reported the SAT as a significant predictor. However, during this same period, and when NCLEX was changed to the pass / fail scoring method, Alexander and Brophy (1997), Beeman and Waterhouse (2001), Dell and Valine (1990), and Wall et al. (1993), all failed to find significance with the SAT as a predictor variable. Crow et al. (2004) (r = -0.4, p <.03) and Haas et al. (2004) (p < .001,

passers exceeded failers in verbal scores) reported significant prediction of SAT scores after NCLEX was administered via computer.

Inconsistent findings with any type of predictor variables raise question as to their value and ability to truly predict NCLEX outcomes. Carpenter and Bailey (1999) concluded that identified predictors to date have not consistently nor reliably predicted NCLEX outcomes. The question was raised as to why faculty evaluate pre-admission variables for their predictive accuracy of NCLEX outcomes. Pre-admission variables assessed 24 to 36 months prior to NCLEX tends to decrease their validity since predictive accuracy of variables is decreased over time (Nunnally & Bernstein, 1994). Pre-admission variables are best preserved for identifying potentially successful students.

Nursing program progressive cognitive variables. Fourteen cognitive measures obtained from nursing theory or clinical course grades were classified as nursing program cognitive variables. Variables included: fourth year / first semester grade, fundamentals theory grade, level one nursing GPA, level two nursing GPA, medical-surgical theory grade, numbers of C's, D's and F's earned in nursing courses, nursing clinical course grade, nursing pharmacology grade, and third year theory course grade, clinical course failure, number of B's or lower in clinical, pediatric nursing course grade, and test average for adult health I nursing course. No significance was found for the four later variables measured as predictors for the computerized NCLEX or for level two nursing GPA during the three NCLEX time periods.

Collins (2002) and Schafer (2002) found level I nursing GPA to be the strongest predictor of NCLEX success during the computerized NCLEX time period, but

Seldomridge and DiBartolo (2004) did not. Fowles (1992) (r = .74,p = .05) found level I nursing GPA to be strongly correlated to the original NCLEX prior to 1988. However Alexander and Brophy (1997), Heupel (1994) and Milan (1997) did not find significance with this variable.

In terms of nursing specialty courses, Gallagher et al. (2001) found the fundamentals theory course to be predictive of success (p <.05). Collins (2002) reported pharmacology grades to be the strongest predictor of NCLEX success (p <.000). Nnedu (2002) (p not provided) and Barkley et al. (1998) (r = 58, p = .001) found the psychiatric nursing course grades to be most predictive. Barkley et al. (1998) (r = .58, p = .001) also reported their school's pediatric course as having a significant correlation, but Nnedu (2000) reported no significant correlation for pediatrics.

Fourth year- first semester course grades' correlation with NCLEX outcomes was evaluated by four different researcher studies (Beeman & Waterhouse, 2001; Beeson & Kissling, 2001; Heupel, 1994; Waterhouse, Carrol, & Beeman, 1993) during the three NCLEX phases, but this variable was found to be significant (r = .24, p < .05) only by Heupel (1994) during the second NCLEX phase (non-computerized pass/fail).

Twelve studies evaluated the medical-surgical theory course grades during the first, second, third and fourth years of nursing school for predictability or correlation with NCLEX outcomes. Horns et al. (1991), Mills et al. (1992a), Drake and Michael (1995), Endres (1997), Alexander and Brophy (1997), Beeson and Kissling (2001) and Ostrye (2001) did not find grades for nursing theory predictive of NCLEX outcomes. However, Lengacher and Keller (1990) found two nursing theory courses to have significant correlations (r = .77, r = .79; p < .05) with NCLEX scores. Heupel (1994), also found

three nursing course grades had significant correlations (r = .44, r = .51, and r = .60; p < .0001), but reported no correlation with fourth year / first semester theory grades. In 1993, Waterhouse et al. reported the fourth year / first semester theory grade significantly correlated (r =.24, p<.05) with NCLEX outcome, but that the third year theory course grade did not. Beeman and Waterhouse (2001) reported that several nursing theory courses contributed to prediction of the computerized NCLEX (r = .30 – r = .38; no p provided). Seldomridge and DiBartolo (2004) found student test averages of the adult health II course correlated (r = .307; p < .0005) with NCLEX success.

Number of C's, D's and F's earned in nursing courses was a predictor variable found to be significant by two researchers. Beeman and Waterhouse (2001) reported that NCLEX failure significantly correlated (r = .39) with students who earned more than three C's in nursing theory courses; and Seldomridge and DiBartolo (2004) reported that the total number of C's earned during the nursing program significantly correlated (r = -0.34; p < .0005). Endres (1997) found that students who earned a D or F grade in nursing theory were more likely (p<.05) to fail NCLEX. Barkley et al. (1998) also studied this variable but found no significance.

Lengacher and Keller (1990) (r = .79, p <.001) and McClelland et al. (1992) (r = .55, p <.001) found that clinical course failures significantly correlated with NCLEX failure. However, Horns et al. (1991), Beeman and Waterhouse (2001), and Drake and Michael (1995) reported no significant correlation with this variable on NCLEX outcomes.

For each of the variables relating to student performance during the nursing program, there were some studies that reported significance and some that did not. The conflicting findings are expected since there are as many differences in curriculum,

teaching and evaluation methods, faculty, and students, as there are studies being done to determine significant evaluation. Faculty from each program seeks a set of predictor variables specific for their curriculum. Despite the many publications regarding significance with nursing program predictors, very few are generalizable to other programs, and it would be safe to assume that a set of variables found predictive of one group of students from a particular program, may not even be able to consistently predict future groups of students from that same program.

Predictive accuracy is more than just identifying individual or combinations of predictors and finding significance one, or even two times. Rather, establishment of predictive accuracy is compared to the construction of a brick wall, one brick at a time. Each significant finding is one brick in the wall of predictive accuracy for that variable. Nursing faculty search for stable predictor variables prove their ability over time to consistently predict NCLEX outcomes. This search for predictors is a challenge in itself, since along with changes in health care, technology, patient care standards, curriculum, students, and faculty, the NCLEX exam itself changes. Every four years a new job task analysis is completed by the NCSBN, the results of which facilitate the design of the NCLEX blue print. As health care changes, the NCLEX will also change. These changes imply that the predictor variable, to be consistently accurate would have to evolve and change too—in a pattern consistent with changes in the NCLEX exam.

Nursing program exit variables. Variables predicting NCLEX success as nursing students near program completion are classified as nursing program exit variables. These include: exit GPA (cumulative college GPA and cumulative nursing GPA) and standardized testing [NCLEX-RN Risk Appraisal Instrument, Mosby Assess Test, NLN

Achievement Test, and Health Education Systems Incorporated (HESI) Exit Exam (E^2)]. Exit variables have been found to be more consistently predictive of NCLEX outcomes because of their standardized nature and because they are designed to measure overall ability. These variables are also more predictive because they are employed closer to the time of NCLEX testing.

Cumulative college GPA was studied by 16 researchers and found to be a significant predictor of NCLEX success 13 times. During the initial NCLEX time frame with issue of individual scores, Lengacher and Keller (1990) (r = .71, p = .01), Foti and DeYoung (1991) (r = .59, p < .0001), Horns et al. (1991) (r not provided, p < .01), Poorman and Martin (1991) (r = .42, p < .05) and Heupel (1990) (r = .60, p < .0001) found that cumulative college GPA was significantly predictive of NCLEX outcomes. When NCLEX changed to the pass / fail format, five researchers including Dell and Valine (1990) (58% of variance, p < .001), Mills et al. (1992) (75% of variance, p < .05), Mills et al. (1992a) (variance not provided, p < .05), Waterhouse et al. (1993) (r = .25, p < .05), Milan (1997), (predicted 86.40% of subjects when combined with NLN med-surg, no p provided), and Arathurzik and Aber (1998) (r = .275, p = .05) found cumulative college GPA to be a significant predictor, but two researchers including Wall et al. (1993) and Endres (1997) did not. Since NCLEX changed to the computerized pass / fail format in 1994 only three research studies have evaluated cumulative college GPA for its predictive accuracy including Lamm (2000) (r = .43, p < .001) and Ostrye (2001) (β= 3.07, p < .02), who found it to be significantly predictive of NCLEX outcome and Haas (2004) who did not. The research from all three time periods support the use of

cumulative college GPA as a consistent predictor, although correlations ranging from .25 - .71 tend to reveal a large range of variability in terms of stability.

Cumulative nursing exit GPA was evaluated for predictive ability in eleven studies and found to be predictive in nine. During the initial NCLEX time frame with issue of individual scores Schaal (1990) (r = .55, p < .001), Foti and DeYoung (1991) (r = .59, p < .0001), and McClelland et al. (1992) (r = .53, p < .001) found nursing exit GPA as a significant predictor variable. When NCLEX changed to the pass / fail format Wall et al., (1993) (r not provided, p < .001), Drake and Michael (1995) (r = .32, p < .001), Alexander and Brophy (1997) (r not provided, p < .001), and Endres (1997) (r =.22, p < .001) also found nursing exit GPA to significantly predict NCLEX outcome. The research for two time periods of NCLEX testing revealed that nursing exit GPA correlations range from .22 - .59, indicating that they may not be as reliable as cumulative college GPA in predicting NCLEX success. Lower correlations with this variable may again be attributed to differences in nursing curriculum, faculty, students, evaluation methods and a variety of other confounding variables.

Since NCLEX changed to the computerized pass / fail format in 1994 four studies have been conducted evaluating nursing exit GPA as a predictor variable. Beeson and Kissling (2001) and Seldomridge and DiBartolo (2004) did not find it to be a significant predictor of NCLEX success, but Giddens (2002) reported that prediction by critical thinking measures including the California Critical Thinking Disposition Inventory (CCDI) or the California Critical Thinking Skills Test (CCTST) were not any more predictive than nursing exit GPA. Giddens (2002) stated there was no need to use the CCDI or the

CCTST, when the nursing exit GPA was just as reliable. Haas (2004) found that NCLEX passers had 0.3 nursing exit GPA higher than NCLEX failers ($p < .0005$).

Nursing programs in search of predictor variables for NCLEX outcomes often times rely on standardized nursing exams. Barkley et al., in 1998, introduced a new tool called the NCLEX-RN Risk Appraisal Instrument (RAI) proposing that a score of less than one significantly correlated ($r = .78$) with NCLEX outcomes. In 2003, Waterhouse and Beeman adapted the RAI for use at their Delaware school and called it the Delaware RAI and tested it for its predictive abilities. This study did not find the Delaware RAI ($r = -.315$) significantly better than a single course grade ($r = -.261$), and reported that their tool correctly classified 74.1% success on NCLEX compared to the 81.6% who were classified for NCLEX success by chance alone. No other studies have mentioned the RAI as a predictor variable. This tool is not considered reliable (Barkley et al., 1998; Waterhouse & Beeman, 2003). The literature reports however, that there are three standardized tests that are used by nursing programs nationwide to predict NCLEX outcomes including the Mosby Assess Test, NLN achievement tests and the HESI E^2.

The Mosby Assess Test was tested seven times during all three time periods of the NCLEX exam and found to correlate with NCLEX outcomes or significantly predict NCLEX outcomes in combination with other variables. During the initial NCLEX time frame with issue of individual scores, Foti and DeYoung (1991) ($r = .66$, $p < .0001$), Fowles (1992) ($r = .79$, $p = .05$) and McClelland (1992) ($r = .66$, $p < .001$) reported the Mosby Assess Test to be a significant predictor, but Wall et al. (1993) did not find it to be so. When NCLEX changed to the pass / fail format, Endres (1997) ($r = .49$, $p < .05$)

found significance and reported students who scored within or below the 21st percentile on the Mosby Assess Test were more likely to fail NCLEX. Since NCLEX changed to the computerized pass / fail format in 1994, the Mosby Assess Test has been studied two times. Beeson and Kissling (2001) found it to be significantly accurate (predicts success 66.7%, $p < .001$) when combined with number of C grades and age. Daley et al. (2003) reported the combined pass / fail predictive efficiency to be 60% ($p < .001$). From 1991 – 2003 Mosby Assess Test correlations with NCLEX outcomes, ranging between .49 - .79, especially when used in combination with other variables, has been a consistent and fairly strong predictor of NCLEX success.

The NLN Achievement Tests are another set of standardized exams that have had fairly good predictive accuracy during the three timeframes of NCLEX testing, especially in comparison to the other three groups of exit predictive variables mentioned. Before 1988, when NCLEX issued individual scores, Lengacher (1990) ($p = .01$) reported the NLN Basics II exam ($r = .66$) and the NLN Psychiatric exam ($r = .70$) to be significantly correlated with NCLEX results. Foti and DeYoung (1991) ($r = .51$, $p < .0001$) and Horns et al. (1991) (r not provided, $p < .01$) also found NLN Achievement tests to be significantly predictive. During the 1988 – 1994 NCLEX time period when NCLEX results were reported as pass / fail, Wall et al. (1993) (r not provided, $p < .001$) , Alexander and Brophy (1997) (80.63% variance when combined with three nursing courses, $p < .0001$), and Milan (1999) (86.40% of subjects pass / fail performance predicted accurately when combined with nursing exit GPA, p *not provided*) found the NLN Exam to predict NCLEX outcomes with moderate accuracy. When NCLEX changed to the computerized format, NLN Achievement tests were again evaluated for

predictive accuracy. Barkley et al. (1998) (r =.59, p < .001), and Briscoe and Anema (1999) (NLN I exam r = .48, NLN II exam r = .37, p = .01) reported NLN exams to be predictive of outcomes for the computerized NCLEX. Crow et al. (2004) reported that the NLN Mental Health (r = .55, p = .02) and Community Health (r = .76, p = .02) exams were predictive of NCLEX. Seldomridge and DiBartolo (2004) reported a correlation of r = .45 for the NLN Comprehensive Achievement Test for Baccalaureate Students (NLNCATBS), demonstrating that when combined with pathophysiology grades, this test predicted 93.3% of NCLEX success and 50% of NCLEX failures for their nursing program. The research regarding the NLN Achievement tests' correlations have ranged from .37 - .76 and have predicted between 80.63% to 93.3% of NCLEX outcomes, with the most recent studies showing the most promise for one data sample. The trend shows that different NLN tests are predictive for different populations of students. In spite of this, the results, to be fully trusted for reliability need to be duplicated and generalizable to a greater population.

Linn and Gronlund (2000) found that the reliability of teacher-made tests usually varies between .60 and .85. Tests of more than 50 items should have reliability coefficients of greater than .80 (Kehoe, 1995). Frisbe (1988) suggests that teacher-made tests tend to yield reliability coefficients that average about .50 and that .85 is the generally acceptable minimum reliability standard when decisions are being made about individuals based on a single score. Correlations ranging from 49 - .79 (Mosby) and 37 - .76 (NLN) demonstrate that the Mosby Assess Test and the NLN Achievement Tests, which are national standardized exams are minimally reliable and certainly should not be used to make decisions regarding student readiness for NCLEX success.

It is important to note that the Mosby Assess Test and the NLN Achievement Tests are pencil and paper exams and the NCLEX is now computerized. The computerized NCLEX format, along with the changes made every four years to the NCLEX exam blue print and passing standard, based on results of the job task analysis, may be contributing factors to these two exam's limitations in predicting NCLEX success. More studies are needed to establish NCLEX predictive accuracy of the Mosby Assess Test and the NLN Achievement Exams.

The last and final predictor variable found in the literature under the exit variable category is the HESI E^2. With this computerized comprehensive exam, (similar to that of the NCLEX computerized format) being relatively new to the nursing arena, literature regarding the predictive accuracy of the E^2 was not published until 1999. Lauchner et al. (1999) evaluated the predictive accuracy of 2,809 RN and PN graduates from the 1996 – 1997 academic year and using a Chi-square analysis found the exam to be 99.49% (p = .05) accurate in predicting NCLEX-RN success when the exam was administered in a monitored setting, and only 96.82% (p = .05) accurate if the exam was administered to students in an unmonitored setting. The total predictive accuracy for this study was 97.41% (p = .05).

Later in 1999, Hanks (1999) wrote a letter to the editor of *Computers in Nursing* questioning the claims Lauchner et al.'s (1999b) results of the E^2's predictive accuracy. Hank's argument posed that predictive accuracy should be calculated based on sensitivity, specificity, positive predictive value and negative predictive value. Lauchner, Newman and Britt (1999a) responded in the same journal in a letter to the editor explaining that using the model proposed by Hank (1999), that the E^2 was 91.15%

accurate in predicting NCLEX-RN failure, referred to as the instrument's sensitivity, and also stated that when the exam was proctored, its ability to predict failures increased to 96.42%. Lauchner et al., (1999a) also states that the positive predictive value and the specificity indicated that the E^2 erred on the side of caution. Students who are identified as having weaknesses in certain content areas are given time to remediate these areas of weakness, and also may be more motivated in their preparations for the actual NCLEX-RN exam.

In 2000, a second study was published reporting the E^2's predictive accuracy of NCLEX success of nursing graduates during the 1997 – 1998 academic year. Newman et al. (2000) analyzed the E^2 scores of 4,042 graduates and determined the exam was 98.74% (p = .05) predictive of NCLEX success. Newman also reported that significantly more low-scoring students failed NCLEX than did high-scoring students.

The following year, a third study by Nibert and Young (2001) reported the predictive accuracy of the E^2 using 6,560 RN and PN graduates during the 1998 – 1999 academic year. Again, the results proved the E^2 to be 97.78% accurate in predicting NCLEX success. Significantly more low-scoring students failed the NCLEX than did high scoring students.

A fourth study was also done (Nibert et al., 2002) for the graduates during the 1999 – 2000 academic year. This study reported a reliability coefficient of .75 for the RN group and .79 for the practical nursing group, averages comparable to those reported in previous year's studies. This study found the E^2 to be 98.46% accurate in predicting NCLEX success.

For four consecutive years, this exam has consistently demonstrated consistent reliability coefficients and high degrees of predictive accuracy for NCLEX success. Compared to the other studies seeking for predictor variables on NCLEX for the past forty years, the E^2 has been the most consistently reliable as well as the most highly predictive. Two reasons the HESI E^2 is predictive is that its items are closely patterned after the NCLEX blue-print, and because it is administered via computer like the NCLEX. Morrison, Adamson, Nibert and Hsia (2004) proposed that four validity studies provide sufficient scientific data to reassure nurse educators that the E^2 can be used confidently to assess students' preparedness for the licensure exam.

Use of E^2 in Nursing Education –Benchmarking and Progression Policies

Based on the aggregate data collected from 19,554 subjects over four consecutive years and the establishment that the E^2 is highly predictive of NCLEX success (Lauchner et al., 1999b; Newman, Britt, & Lauchner, 2000; Nibert & Young, 2001; Nibert, Young, & Adamson, 2002), increasingly more schools of nursing are choosing to administer it to assess student competency and evaluate achievement of curricular outcomes (Morrison et al., 2004; Morrison, Free, & Newman, 2002; Nibert, 2003). According to the HESI database records, nursing programs using HESI exams increased from 85 in December 1999 to 565 in December 2003, an increase of 565% in four years (HESI, 2003).

Developers of the E^2 and users alike claim the exam's main purpose is to assess student's readiness for the NCLEX (Morrison et al., 2004), pinpointing student's subject content weaknesses to provide an invaluable asset in designing individualized remediation programs (Engelmann & Caputi, 1999; Morrison et al., 2002). Nursing

faculties are held accountable for NCLEX pass rates and desire to maintain acceptable pass rate standards by providing remediation services for students who are identified as at risk prior to those students taking the NCLEX (Morrison et al., 2002). In schools that use the E^2 for remediation, significantly fewer low-scoring students have been found to fail the licensure exam with faculty attributing this finding to student participation in remediation (Newman et al., 2000).

Some schools of nursing are implementing progression or remediation policies based on E^2 scores as a means of assessing student readiness for NCLEX, implementing remediation where indicated, and ultimately improving NCLEX pass rates (Morrison et al., 2002; Nibert et al., 2002). Benchmarking for progression is not only a new trend in nursing education, but one that may signify a turning point in higher education (Nibert, 2003). Benchmarking involves designating a specific E^2 score as a minimally-acceptable score students are required to attain (Nibert, 2003).

In a study by Morrison et al. (2002), nursing education administrators of seven RN programs were surveyed about their use of progression and remediation policies and their corresponding NCLEX-RN pass rates. Morrison et al. (2002) found that within two years of implementing progression and remediation policies, NCLEX-RN pass rates had increased by 9% to 41%, concluding that progression polices were highly effective in increasing NCLEX pass rates.

In a HESI Newsletter faculty were encouraged to implement progression policies as a means for external curriculum evaluation (Morrison, 2000). The recommendations for these types of policies included testing nursing students in their last semester of the program using the E^2; having a pre-identified minimum E^2 score that students needed to

obtain, and not allowing those students who didn't obtain the designated benchmark to take the NCLEX-RN (Morrison, 2000). According to the newsletter, students should be allowed to test a total of three successive times if they do not achieve a minimum score on the E^2. The first test should come at the beginning of the last semester, the second should come six weeks before expected graduation, and the third test should come in the last two weeks of the curriculum (Morrison, 2000)

Nibert (2003), desiring to provide faculty with further evidence as to use of the E^2 in progression policies, surveyed 158 nursing programs who utilized the E^2 during the 1999 – 2000 academic year asking programs regarding their use E^2 as a benchmark for progression and remediation. Results from 149 responding RN programs found forty-five (30.20%) schools had either implemented or maintained progression policies during the survey year (Nibert, 2003). Review of policies showed that one or more consequences were most often cited for students who did not achieve the benchmark E^2 score designated by the school, including: 1) denial of eligibility for graduation; 2) an incomplete or failing grade in the capstone course; and/or 3) withholding of approval for NCLEX candidacy (Nibert, 2003). Survey results also found that forty (88.9%) of the forty-five schools had adopted policies enforcing mandatory re-testing using a different version of the E^2 for students that failed to achieve their schools' specified E^2 benchmarks (Nibert, 2003). Thirty-six of the forty-five RN programs required retesting if the students did not achieve the benchmark score on the E^2 (Nibert, 2003).

Nibert (2003) also surveyed nursing programs for the level of benchmark identified as the minimally acceptable score designated by the progression policies. Of the forty-five policies, thirty-four (75.56%) designated an E^2 of 85 as the required

benchmark for progression. Only two (4.44%) reported using an E^2 score below 85, and six (13.33%) indicated that they used an E^2 score higher than 85, with 90 being the highest reported score used as a benchmark (Nibert, 2003).

Nibert's (2003) study investigated whether or not program progression policies required remediation if the student failed to achieve the minimally acceptable E^2 score. Of the 149 responding program administrators, 107 (71.81%) reported that remediation was not required, but that optional remediation was offered. The remaining forty-two (28.19%) schools required remediation based on E^2 score which entailed attendance at a designated remediation course, completion of software programs, comprehensive review guided by NCLEX preparation books, and/or mandatory tutoring sessions with faculty (Nibert, 2003).

Schools reported that implementation of a progression policy based on E^2 scores was no easy task (Morrison et al., 2002). This difficulty may be because of the associated consequences tied to the student E^2 scores. It is clear that since nurse educators desire to maintain or improve pass rates, they may be considering implementation of a progression policy. Faculty may consider a number of factors as they develop a progression policy specific to their nursing program and student needs.

First, there must be quantitative evidence that the E^2 continues to be highly predictive of NCLEX success, and also data presenting the degree of risk for failing the NCLEX associated with various E^2 scoring intervals. Second, the level of the E^2 benchmark needs to be considered. Nibert (2003) recommended that in addition to what is currently being done in relation to E^2 score benchmarking in other nursing programs, faculty also carefully consider the characteristics of their programs (i.e., current pass

rates and size of the program), and things that could influence the decision. Third, faculty need to consider the type and numbers of consequences that will be best for their student population and program objectives. This study attempts to re-establish the predictive accuracy of the E^2 for an additional academic year, utilizing the E^2's new scoring method based on difficulty of individually weighted items, and will examine student NCLEX performance associated with various E^2 scoring intervals. This study will also examine the impact that progression policies may have on E^2 student performance. Specifically, the research is focused on the impact of the designated level of the benchmark and the type and numbers of consequences progression policies designate are associated with the E^2.

Establishment of consequences in relation to the E^2 may change student perception of the exam, and additionally, may change the students' approach toward taking the exam. Whereas students who have no consequences attached to the E^2 may perceive the exam as "just another nursing exam," students who attend schools that have implemented consequences associated with the exam may tend to perceive this exam as a "high stakes" exam similar to NCLEX.

Motivation and Consequences in Testing

Academic achievement, learning, and NCLEX performance are major constructs of interest to nurse educators and educational researchers in nursing. Many schools of nursing and a handful of health care institutions have come to rely on the established ability of the E^2 to predict NCLEX outcomes. The literature shows significant variance in student performance on standardized examinations is related to motivation levels of students (Burke, 1991; Schmidt & Ryan, 1992; Smith & Smith, 2002; Sundre, 1997b,

1999; Sundre & Kitsantas, 2004; Wolf & Smith, 1995; Wolf, Smith, & Birnbaum, 1995). This variance often occurs in the absence of consequences for examinees (Sundre, 1999). Apparently, students having the knowledge necessary to successfully complete an exam is not sufficient; an affective component also effects the testing outcome (Pintrich & DeGroot, 1990).

For some students taking the E^2, their test score is used to calculate a portion of the capstone course final grade or is tied to a progression policy for graduation or required remediation (Morrison et al., 2002; Nibert et al., 2002; Nibert, Young, & Britt, 2003). These students are being tested under consequential test conditions and it is assumed that they are highly motivated to perform well (Sundre & Kitsantas, 2004). Other students take the E^2 with no consequence attached to their score. In this situation, students are being tested in conditions for which no consequences or stakes are present. Non-consequential or "low stakes" test conditions are those for which the test results have no significant bearing for the students taking the test, however the test results are used for individual study, internal curriculum evaluation and as self study outcome criteria for accreditation organizations such as NLNAC or CCNE (AACN, 2002; Gabbin, 2002; Morrison, 2000; Morrison et al., 2004; Morrison et al., 2002; Nibert et al., 2003; NLNAC, 2004; Olsen & Wilson, 1991; Sundre, 1997a, 1999; Wyatt, 2002).

There is substantial evidence indicating that the disposition of test-takers is central to performance (Schmidt & Ryan, 1992). Research has demonstrated the positive impact of test-taking motivation in consequential conditions and the negative impact of low motivation in non-consequential conditions (Banta & Pike, 1989; Jakwerth

& Stancavage, 2003; Napoli & Raymond, 2004; Olsen & Wilson, 1991; Sundre, 1999; Sundre & Kitsantas, 2004; Wolf & Smith, 1995; Wolf et al., 1995).

Napoli and Raymond (2004) tested two groups of college students (graded condition & un-graded condition) enrolled in an introductory psychology course using a twenty-item multiple-choice assessment instrument (PsyOA) developed for use in introductory psychology classes for the purpose of General Education assessment by a committee of psychology faculty. Both groups had identical syllabi, textbooks, lectures and assessments throughout the course, and both groups had very similar demographics of high school average, overall college GPA, age, gender and College Placement Test scores. Despite similarities in all respects, at the conclusion of the course, when the twenty-item exam was administered to both groups of students, test results under graded conditions (n = 46) were significantly higher [t (78) = 5.62, p < .001)] and more reliable (r = .71) than results from the test administered to students in the non-graded condition (n = 34; r = .29). Students responded to testing situations quite differently based on the circumstances surrounding motivation and consequences.

Olsen and Wilson (1991) used personal follow-up interviews with 'low' scorers (n = 45) and identified decreased motivation as a key factor in explaining lower than expected College Outcome Measures Project (COMP) scores among college sophomores. Test results were perceived as being of little importance or relevance, and students identified lack of motivation to perform well on their sophomore examination as the primary reason for the low scores (Olsen & Wilson, 1991).

Banta and Pike (1989) interviewed college seniors (N = 3,485) following their completion of either the 60 item College Outcome Measures Project (COMP) or the 144

item Academic Profile from Educational Testing Service (ETS). Students were randomly assigned to either of the two testing conditions when they reported for college exit testing, a mandated graduation requirement at the University of Tennessee, Knoxville (UTK). Interview results showed students self reported low effort on these mandated higher education assessment activities and authors attributed this to having no direct consequences tied to student scores (Banta & Pike, 1989).

Jakwerth and Stancavage (2003) interviewed eighth grade students (N = 65) in three states after completion of the National Assessment of Educational Progress (NAEP) exam. In regards to motivation of students in tests without consequences, Jakwerth and Stancavage (2003) reported that for low achievers, motivation was a pervasive problem.

Wolf, Smith, and Birnbaum, (1995) explored the impact of task difficulty on motivation in test-taking by administering a math exam to two groups of high school students. One group (n = 133) had no consequence attached to performance, and the second group (n = 168) were told test scores would determine placement into required remedial math programs (Wolf et al., 1995). The instrument was a 30-item multiple choice math test consisting of numerical operations, measurement and geometry, patterns and functions, data analysis, and fundamentals of algebra. Items were characterized by item difficulty (from p values), degree of mental taxation (how much mental effort was necessary to reach a correct answer), and item position (as an index of the level of fatigue of the test taker) (Wolf et al., 1995). Following testing, students were asked to respond to a list of attitudinal questions about the exam, one of which was

"I worked hard to answer the questions on this test" A. strongly agree; B. agree, C. disagree, D. strongly disagree (Wolf et al., 1995).

The exam for the consequence group had a reliability of .80, and exam reliability for the no consequence group was .84 (Wolf et al., 1995). The overall mean performance for the group with consequences (m = 15.56) was not statistically better [t (299) = 0.735; p not provided] than the performance of the group without consequences (m = 15.07), however, there was a significantly higher motivation expressed from the group with consequences attached compared to the group without consequences [t (284) = 3.79, $p < .001$) (Wolf et al., 1995). Lower levels of motivation did not affect reliability of the test, but regression analysis of motivation, difficulty, degree of mental taxation and fatigue showed significant differences in performance between the two groups based on the degree of mental taxation and item difficulty (Wolf et al., 1995). Results indicated that students perform fairly well on straightforward items, but have difficulty with more complex items or items with multiple steps. When items are more mentally taxing, students tend not to give them sufficient effort if the test is non-consequential. Authors concluded that lack of motivation negatively influences test performance above and beyond known ability levels (Wolf et al., 1995).

Wolf and Smith (1995) conducted an experimental study using 158 college undergraduate students enrolled in a child development course responding to two parallel examinations under two experimental conditions. For one of these examinations, the score counted as part of the course grade (consequence), and in the other condition it did not (no consequence). The two forms of a forty-item multiple choice exam covering the first four units of the child development course, were constructed by

grouping items by content area and then assigning them to forms based on item statistics from previous administrations (Wolf & Smith, 1995). Following each test administration, subjects completed an eight-item motivation scale. Test conditions were counter balanced. Wolf and Smith (1995) reported that the examination condition with course credit consequences resulted in significantly higher reported motivation [t (157) = 16.66; $p < .01$;ES = 1.45] and test score performance [F (1, 154) = 15.50, $p < .001$;ES = .26].

Sundre (1999) replicated Wolf and Smith's (1995) research using undergraduate psychology majors (N = 90) and two parallel examinations from a personality psychology course administered under two different testing conditions--consequence (counts toward the grade) and no consequence (doesn't count for the grade). The test instruments, developed by Sundre and a faculty colleague, consisted of thirty multiple choice items and one essay item. Both exams were administered consecutively to subjects using a counterbalanced design and random assignment. Following each section of the test, subjects completed a ten-item motivation questionnaire. Sundre (1999) hypothesized that motivation to perform was influenced by the consequence associated with test performance and the consequential testing condition would lead to a better test performance than a non-consequential testing condition. Similar to Wolf and Smith's findings (1995), Sundre (1999) reported subjects in the consequential testing condition performed better [t (61) = 3.54; p = .001, ES = .62] than students in a non-consequential testing condition.

In 2002, Smith and Smith replicated research performed by Wolf and Smith (1995) and Wolf et al. (1995) in a different setting as further validation that consequence

of a test to an examinee influences test motivation. Smith and Smith (2002) administered two forms of a forty-item exam to four sections of undergraduates students enrolled in educational psychology (N = 112). One exam counted for the student's grade, the other did not. Following each test, participants completed the Posttest Index of Test Motivation (Wolf & Smith, 1995), consisting of eight 5-point items concerning the participant's motivation. Analysis of variance with test score as the dependent variable indicated significance for main effect of consequence [F (1,108) = 32.95, p < .01, ES = .44], demonstrating that consequence of a test to an examinee influences test performance (Smith & Smith, 2002). A second analysis of variance with motivation score as the dependent was also significant for main effect of consequence [F (1,108) = 150.26, p < .01, ES = 1.58], indicating that consequence of a test to an examinee influences motivation of performance.

Weiner's (1986) attributional theory of motivation and emotion provides a clear theoretical basis by which consequential versus non-consequential test conditions can be compared. Namely, non-consequential test conditions present individuals with expectancy, value, and affect considerations that may undermine willingness to perform optimally. Students facing the E^2 in a non-consequential test condition may perceive it as "just another test" or as being minimally important since NCLEX is weeks or months away, possibly believing that there is plenty of time for focused preparation for the "real" test that counts towards earning a nursing license. Also, the E^2 having 160 high level critical thinking questions, may require more effort than senior nursing students with busy schedules may be willing to exert if no consequences are tied to the score.

This study is the first to examine the impact of one or more types of consequences on E^2 performance. Research about student E^2 performance is important because to date, the major participant in assessment activities related to the E^2 has been largely ignored. The psychology of the examinee is a critical factor in all assessment results and may assist faculty in determining whether or not a progression policy is indicated according to nursing program objectives and characteristics. Results of this study may help faculty decisions in determining the actual designated level of benchmark, or number of consequences that would be appropriate in beginning or modification of a current progression policy associated with E^2 score.

Summary

The topic of predicting success on the NCLEX has been thoroughly researched for more than 40 years. This review of the literature captured only a span of the last fifteen years in order to evaluate the current state of nursing science for predictive accuracy of NCLEX success using all of the literature published from 1990 to 2005. Forty-two research studies evaluating correlation between 89 predictor variables and NCLEX outcome were analyzed to determine which predictor variables were the most consistently predictive of NCLEX outcome. The sample sizes of the research studies ranged from 38 – 6,800 subjects. None of the 42 studies mentioned calculation of statistical power in determining sample size. Correlation, regression and discriminant analysis were the common approach to identifying significant predictive variables. In 1988, when NCLEX changed to reporting only pass / fail results, predicting NCLEX success became more of a challenge because no longer were there individual NCLEX scores to correlate with other quantitative measures.

Researchers have been unable to consistently and reliably demonstrate NCLEX predictive accuracy using non-cognitive factors, pre-admission factors or nursing program progressive factors. The nursing program exit cognitive variables were found to be the most predictive of NCLEX success, with comprehensive examination used most frequently by nursing faculty. The Mosby Assess Test correlations with NCLEX outcomes ranged between .49 and .79, especially when used in combination with other variables. The NLN Achievement Tests correlations with NCLEX ranged from .37 - .76 predicting between 80.63% and 93.3% of NCLEX outcomes. The HESI Exit Exam had consistently high reliability (.75 - .79) and high predictive accuracies (97.41% to 98.3%) published over four consecutive years.

With teacher-made test reliability ranging between .60 and .85, it is expected that commercial, standardized tests maintain a high standard of reliability, especially if decisions regarding student readiness for NCLEX will be made based on this exam. The literature clearly demonstrates that the HESI E^2 is most predictive of all eighty-nine identified and studied predictor variables. Therefore, the HESI E^2 was chosen as the focus of the current study because it was the only comprehensive nursing exit exam with consistently high predictive accuracies.

Establishment of predictive accuracy is likened unto building a wall of bricks. Each study provides one more brick of evidence needed in establishing the strength of the predictive accuracy of a variable. This study's aim was to analyze the predictive accuracy of the E^2 for graduates who tested NCLEX during the 2001 – 2002 academic year. The HESI E^2 needs annual evaluation to maintain that it is still predictive of nursing graduate NCLEX outcomes throughout the nation. With changes continually

being made to NCLEX to reflect differences in technology and health care practices, student groups, program utilization, as well as curriculum content and standards— also are changes made to the HESI E^2, providing sufficient justification that this exam needs to be evaluated for its predictive abilities regularly. For now, however, the literature review appears to indicate that the HESI E^2 is the consistent predictor variable nursing faculty have sought for almost half a century.

Increased utilization of the HESI E^2 has not only proven effective to guide student remediation, but also improved NCLEX pass rates through the use of progression policies. During the 1999 – 2000 academic year, 45 of 158 (30.20%) schools administering the HESI E^2 indicated use of such a policy and currently more than 50% of schools who administer the E^2 have some type of progression policy in place (Dr. Ainslie Nibert, Director of Research, HESI, personal communication, August 2, 2005).

Progression policies cause the E^2 to be viewed by the students as a high stakes examination, similar to that of the NCLEX. Student performance on the E^2 may be related to motivation level and emotional disposition. Studies have shown the positive impact of test-taking motivation in consequential conditions and negative impact of low motivation in non-consequential conditions where students in consequential conditions were more motivated and performed better compared to non-consequential conditions. Therefore, Weiner's attributional theory of motivation and emotion was chosen to guide this study to evaluate how presence of consequences impacted student performance on the E^2 and determine whether there were differences in student performance based on severity or amount of consequences.

Nursing faculty who currently are using the E^2, but do not have a progression policy in place may be considering establishing a progression policy based on E^2 scores and may desire information not only about the exam's continued predictive accuracy and the degree of risk associated with various E^2 scoring intervals, but also information regarding what level of benchmark would be best to designate and what type(s) of consequences should be associated with student E^2 scores at their schools. The literature described types of progression policies which have been implemented and indicated that resistance from students or faculty may be experienced with progression policy proposal / implementation. School administrators reported that implementation of a progression policy based on E^2 scores was no easy task.

To date, there has been no nursing literature reporting the impact of progression policies on student performance. Describing current use of progression policies by nursing program administrators and evaluating the impact progression policy consequences have on student performance on the E^2 may prove valuable as faculty develop and implement E^2 progression policies.

This study not only evaluated the E^2's predictive accuracy and student NCLEX performance associated with the various E^2 scoring intervals, but also focused on the impact of progression policies on student E^2 performance and how progression policies motivated that performance. This retrospective database study, utilized student E^2 scores, NCLEX outcomes, and program progression policy information to evaluate how designated levels of the E^2 benchmark and the type and number of consequences progression policies designated, were associated with student E^2 performance.

CHAPTER 3

PROCEDURE FOR COLLECTION AND TREATMENT OF DATA

This research utilized a descriptive, retrospective design to 1) assess the accuracy of the E^2 in predicting success on NCLEX based on the HPM's new individual weighted item difficulty scoring method; 2) identify differences in NCLEX outcomes for students whose E^2 scores were in the A/B, C, D, E/F, and G/H scoring intervals; 3) identify differences between mean E^2 scores of students enrolled in nursing programs that have adopted benchmark progression policies and those enrolled in programs that have not adopted such policies; 4) identify differences between mean E^2 scores among programs that specify one of three ranges of a designated benchmark E^2 score as requirement for graduation described as 90, 85 or 80 and below; 5) identify differences between mean E^2 scores among programs that specify one of four ranges of capstone course final grade weights for the E^2 scores described as 1-5%, 6-10%, 11-15%, and 16-20%; 6) identify differences in mean E^2 scores among programs that specify one of three ranges of a designated benchmark E^2 score for required remediation described as 90, 85, or 80 and below; and 7) identify differences in mean E^2 scores among students in nursing programs that use a single benchmark consequence or a combination of two or three benchmarking consequences.

Descriptive studies gather data about variables of interest to explain a problem or situation, identify problems with current practice, justify current practice or make judgments (Burns & Grove, 1997). HESI has a computerized database consisting of

student scores on all HESI examinations, NCLEX outcomes and nursing program progression policy information.

Setting

The HESI database was the source of information regarding student performance. The database contained information from administrators from BSN, ADN, and diploma schools of nursing that administered the E^2 between September 1, 2001 and May 31, 2002 (Year V). During the study year, 277 schools of nursing across the United States with a total enrollment of more than 14,000 students administered the E^2 to students in their final semester of the nursing program. Exams were administered in a monitored setting, and data were returned to HESI for analysis.

In addition to containing student E^2 scores and NCLEX-RN outcomes, the database contained information provided by program administrators describing any progression policies in existence tying student E^2 scores to a specific consequence or number of consequences during the designated study year. Information in the database regarding progression policies and E^2 score benchmarking practices, if any, included whether the E^2 score was used as a weighted portion of the capstone course final grade, and if so, what the specific weight was; and whether a progression policy stipulated a minimum E^2 score as a requirement for graduation, permission to take NCLEX, and/or mandatory remediation. If a required benchmark was stipulated, the specific benchmark score was available.

Population and Sample

The sample for this study was nursing students (N = 14,727) attending associate, baccalaureate or diploma registered nursing programs who took the E^2 in their last

semester of the nursing program during the 2001 – 2002 academic year and whose nursing program administrators had provided NCLEX outcome and progression policy information about these students. The E^2 was administered within the last 6 months of graduation.

Protection of Human Subjects

Because this project was using a database of educational tests and nursing program progression policy information stripped of student and school identifiers, it qualified for exempt review. Approval was granted by the Texas Woman's University Institutional Review Board—Houston Center in September, 2005. (See Appendix B). Additionally, no identifing data regarding students, schools, or administrators was incorporated into this dissertation or published articles of this study's findings.

Instrument

The HESI E^2 is a comprehensive computerized nursing examination that is administered to students in their last semester or quarter of a nursing program. It simulates the NCLEX in that it follows the NCLEX-RN test blueprint developed by the National Council of State Boards of Nursing (NCSBN) (Nibert, Young, & Adamson, 2002).

Test items for the E^2 are developed using a critical thinking model (Morrison & Free, 2001; Morrison, Smith, & Britt, 1996), which requires application of clinical nursing decision making to determine correct responses. Each version of the E^2 is created from test banks containing questions written specifically for HESI by a national team of nurse educators and clinicians.

The HESI Predictability Model (HPM), a proprietary mathematical model, is used to calculate all HESI scores. This computation does not produce a percentage score. Instead, the HESI score reflects application of the mathematical model to raw scores. The HPM considers several factors, including the difficulty level of each test item, which is determined by dividing the number of correct responses to a test item by the number of responses to that same item, and multiplying by one (Crocker & Algina, 1986). Prior to 2001, the average difficulty level of all test items contained in an exam was considered in the calculation of the HPM. Beginning in January 2001, HESI began using the individual difficulty level of each test item in the calculation of the HPM (Susan Morrison, personal communication, January 12, 2003). With the new scoring method, students answering the same number of test items were likely to receive different scores because their scores were not only dependent on the number of test items answered correctly, but also on the difficulty of test items answered correctly. More difficult test items received more weight than less difficult test items. For example, a HESI score of 86 might be a percentage score of 68%, depending on the difficulty level of the individual test items contained on a particular test or in a particular category of a test. The predictive accuracy of the E^2 with weighted scores has yet to be established.

The 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, 2001) three National League for Nursing Accrediting Commission (NLNAC) categories (NLNAC, 2004); and 17 categories described by the American Association of Colleges of Nursing (AACN) (AACN, 2002).

Reliability

An aggregate item analysis is performed each time a school returns student data to the company. 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 (Nibert et al., 2002). The average KR-20 for the E^2 was 0.92 during the 2001 – 2002 academic year.

Linn and Gronlund (2000) found that the reliability of teacher-made tests usually varies between .60 and .85. Tests of more than 50 items should have reliability coefficients of greater than .80 (Kehoe, 1995). Frisbe (1988) suggests that teacher-made tests tend to yield reliability coefficients that average about .50 and that .85 is the generally acceptable minimum reliability standard when decisions are being made about individuals based on a single score.

Validity

Validity of the E^2 is determined by an estimation of content validity, construct validity, and criterion-related validity (Nibert et al., 2002). Content validity refers to the test items' efficiency in measuring students' fundamental nursing knowledge and skills. Proficient nurse educators and clinicians established content validity for the E^2 by evaluating the test items' significance to entry-level practice. Construct validity refers to the extent to which the test measures specified traits or attributes at a conceptual level. As a comprehensive exit examination, the E^2 measures constructs that are essential to entry-level nursing practice as defined by the NCSBN job analysis studies and reflected

in the NCLEX test plans. Criterion-related validity refers to inferences made from analyses of students' E^2 scores for the purposes of predicting NCLEX success. Annual research studies that correlate E^2 scores with actual NCLEX outcomes offer further evidence of the examination's predictive validity.

Data Collection

HESI staff obtained information for the database by sending schools that administered the E^2, an electronic summary analysis of their aggregate data. One of the reports in this summary analysis was a grouping of students' scores by scoring categories. These categories ranged from A/B, the highest-scoring category, to G/H, the lowest scoring category. These scoring categories served as the basis for formulating the scoring intervals to be used for data collection and analysis. The scoring intervals were made up of HESI scores designated as: A, scores 95 or greater; B, scores from 94-94.99; C, scores from 85-89.99; D, scores from 80-84.99; E, scores from 75-79.99; F, scores from 70-74.99; G, scores from 65-69.99 and H, scores $\leq$ 65.

HESI staff also sent school administrators at participating schools of nursing an electronic mail message inviting participation in the study. The invitation contained an electronic link to their student's information on HESI's database. School administrators clicked on the link and were given the list of students scoring in the individual scoring categories. They were asked to identify from the list, the number of students who failed NCLEX.

The second portion of the survey questioned respondents regarding their school's use of the HESI E^2 in relation to a progression policy, asking for identification of specific uses of student E^2 scores in their nursing program. E^2 use options included: 1)

scores used as a diagnostic tool to guide student preparation for NCLEX, but not used as a benchmark for progression or graduation; 2) scores used as a weighted portion of the capstone course grade; 3) scores not meeting a required benchmark as per program policy indicate student ineligibility to progress to graduation and / or to receive permission to take NCLEX; 4) scores not meeting a required benchmark as per program policy indicate requirement of student participation in remediation activities. Response driven follow-up questions were then presented to clarify specifics about the program's progression policies. Responses to the on-line survey were immediately saved in the HESI database. After two weeks, schools failing to respond to the first email were sent a follow-up email. If no response occurred after the second email, each dean/director received a personal telephone call verifying their receipt of the email and encouraging them to participate. Three weeks following the phone contact, a hard copy of the survey was sent to all schools who had still not responded. Each school who responded to the survey was awarded a free 30-day access to an on-line HESI case study of their choice.

Treatment of Data

Upon IRB approval, HESI provided the researcher with aggregate student scores by school, and school listings categorized by program type, region serviced (rural or urban), funding sources (public or private), student to faculty ratios, NCLEX pass rates and program progression policy practices. HESI removed school and student names from the data provided to assure confidentiality of school and student information. (See Appendix C). The researcher consolidated and analyzed the survey results data for accuracy and for descriptive qualities such as frequencies, mean, median, mode and tabular results.

Predictive accuracy, research question 1, was determined using the most stringent method, by examining only the NCLEX outcomes of those who were predicted to pass, which consist of those scoring in the categories A or B. The number of students scoring in category A/B who failed NCLEX was divided by the total number predicted to pass and subtracted from one. Chi-square goodness of fit tests were used to detect differences between expected and observed frequencies among NCLEX outcomes of students scoring at each of the five HESI scoring intervals (research question 2).

To answer the third research question, an independent *t*-test was used to evaluate differences between mean E^2 scores of students enrolled in nursing programs that had adopted benchmark progression policies and those enrolled in programs that had not adopted such policies. The fourth, fifth, sixth and seventh research questions were answered using one-way between subjects analysis of variance testing (ANOVAs) to determine differences in mean E^2 scores among programs that specify: 4th) one of three ranges of a designated benchmark E^2 score as requirement for graduation described as 90, 85 or 80 and below; 5th) one of four ranges of capstone course final grade weights for the E^2 scores described as 1-5%, 6-10%, 11-15%, and 16-20%; 6th) one of three ranges of a designated benchmark E^2 score for required remediation described as 90, 85, or 80 and below; and 7th) use of a single benchmark consequence, or a combination of two or three benchmarking consequences. If any of the main ANOVA's were significant, post-hoc pairwise follow up testing evaluated differences between groups. For each of the four main ANOVA analyses, alpha was adjusted using Bonferonni technique and set at .0125 to prevent a type I error.

CHAPTER 4

ANALYSIS OF DATA

A descriptive retrospective design was used to examine the predictive validity of the Health Education System Incorporated (HESI) Exit Exam (E^2) in predicting National Council of State Boards of Nursing Examination for Registered Nurses (NCLEX-RN) success. This study also explored whether or not students perceive the HESI E^2 as a high stakes exam on the basis of having an existing benchmark or consequence associated with their final score; and whether or not this perception impacted their performance on the exam as reflected by their score. Descriptive studies are done to gather data about variables of interest to explain a problem or situation, identify problems with current practice, justify current practice or make judgments (Burns & Grove, 1997). This chapter presents a descriptive analysis of the sample and the findings for the seven research questions.

Data regarding E^2 scores, NCLEX outcomes and progression policies tied to E^2 scores were obtained from the HESI database. Previous to this study, HESI staff obtained information from administrators at schools of nursing that administered the E^2 during the academic year 2001 – 2002. This chapter addresses the description of the Year V sample and the findings generated from data analysis.

Description of the Sample

Of the 277 program administrators invited to become part of the HESI database, 182 (65.7%) submitted completed electronic questionnaires describing E^2 testing of 9,695 students. Seventy-one (39%) program administrators were from BSN programs,

contributing 3,668 (37.8%) student scores. One hundred three (57%) program administrators were from ADN programs, contributing 5,707 (58.9%) student scores. Eight (4%) program administrators were from diploma programs, contributing 320 (3.3%) student scores. Table 3 shows the breakdown of the RN sample by type of program.

Table 3.

RN Sample Breakdown by Type of Program

Description	ADN	BSN	Diploma	Total
Number	5707	3668	320	9695
Percentage	58.9%	37.8%	3.3%	100%

Findings

Predictive Accuracy of the E^2 Regarding NCLEX Success

The first research question addressed the predictive accuracy of the Health Education Systems Incorporated (HESI) Exit Exam (E^2) using the weighted scoring method based on individual item difficulty for students in baccalaureate, associate degree, and diploma registered nursing (RN) programs during the 2001 – 2002 academic year. Predictive accuracy was determined by examining only the NCLEX outcomes of those who were predicted to pass, which consist of those scoring in the categories A or B. The number of students scoring in category A/B who failed NCLEX was divided by the total number predicted to pass and subtracted from one.

In Year V, there were a total of 8,009 nursing students with scores in the A/B category, 3,251 (40.6%) BS students, 4,478 (55.9%) AD students, and 280 (3.5%) diploma students. This figure is less than that of the original sample because students

whose HESI scores were in the database were dropped for this portion of the data analysis if NCLEX outcomes were not provided for them by the school administrators. NCLEX outcomes may not have been provided for these missing students if these students received a failing grade in the capstone course, were not permitted to progress to take NCLEX, or did not choose to take the NCLEX prior to data collection.

Scores in the A/B category indicate that students are predicted to pass the NCLEX-RN without additional preparation. Of the 2,714 students scoring in the A/B category, 1,508 (55.6%) were from AD programs, 1,124 (41.4%) were from BS programs, and 82 (3.0%) were from diploma programs. Only 59 (2.2%) of those students predicted to pass by scoring in the A/B category actually failed NCLEX. Twenty-three (38.9%) failures were AD students, 33 (55.9%) failures were BS students and three (5.1%) failures were diploma students.

For the 2001 – 2002 academic year, the overall predictive accuracy of the E^2 for all students was 97.8%. The predictive accuracy of the E^2 for BS student performance on NCLEX was 97%. For AD students it was 98.4%, and for diploma students it was 96.2%.

Differences in NCLEX Outcomes by HESI Scoring Intervals

The second research question asked if there was a difference in NCLEX outcomes for students whose E^2 scores were in the A/B, C, D, E/F, and G/H scoring intervals. NCLEX-RN outcomes for the five HESI scoring intervals were compared using a weighted chi-square analysis of the student's scores for each of the five scoring intervals. The analysis revealed significant differences among student scoring intervals [χ^2 (4, N = 8009) = 801.51, p < .0005].

Analysis of scoring interval data indicated that NCLEX-RN failures increased as the scoring interval decreased. Of the 2,714 students who scored in the A/B category (E^2 score of 90 or greater), 59 (0.9%) failed the licensing exam; of the 1,300 students who scored in the C category (E^2 score of 85 - 89.99), 52 (4.0%) failed; of the 1,263 students who scored in the D category (E^2 score of 80 – 84.99), 85 (6.7%) failed; of the 1,816 students who scored in the E/F category (E^2 score of 70 – 79.99), 267 (14.7%) failed, and of the 916 students who scored in the G/H category (E^2 score of less than 69.99), 287 (31.3%) failed. Figure 2 illustrates the patterns of NCLEX-RN 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 AD students were compared with A/B interval scores of BS students. NCLEX outcomes of students attending diploma programs were not included in this comparison because three of the cells contained less than five subjects.

No significant differences were found in the NCLEX outcomes between the two program types (AD and BS) for students scoring at like intervals in each of the five scoring intervals. In summary, significant differences in NCLEX-RN pass rates were noted for each scoring interval, but no significant difference was found between AD and BS programs when they were compared at like intervals.

Student E^2 Performance with and Without Progression Policy Consequences

The third research question addressed whether nursing students perform differently on the E^2 in the presence of consequences as mandated by progression policies. Exit Exam scores and frequency data regarding existence of a consequence

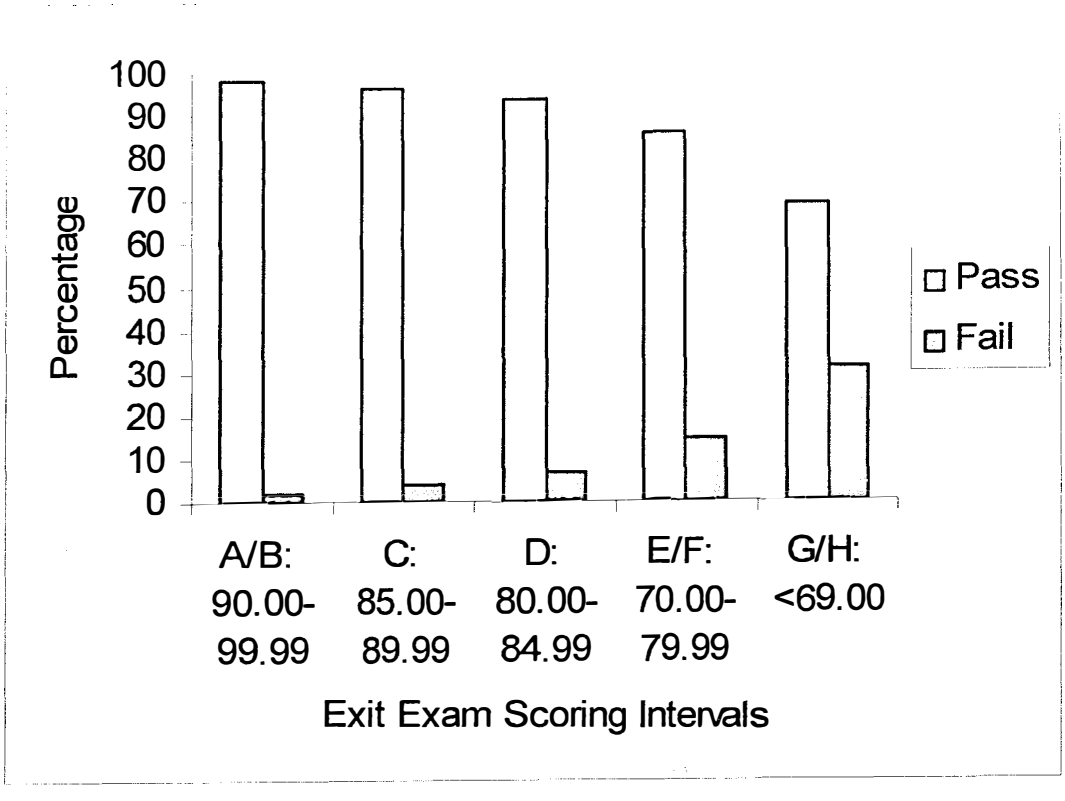

Figure 2. NCLEX-RN pass/fail rates by E^2 scoring intervals.

tied to E^2 score were obtained from electronic and hard copy questionnaires sent to participating schools. Administrators were asked to select one to four statements that best described their school's use of E^2 scores including: 1) Used as a diagnostic tool to guide students' preparation for NCLEX—but not used as a benchmark for progression or graduation; 2) Used as a weighted portion of the capstone course grade; 3) Used for progression to graduation and/or permission to take the NCLEX; and / or 4) Used to require remediation for those students who did not achieve a designated benchmark E^2 score.

Responses of participating program administrators were analyzed to determine how the E^2 was used within their programs. Of the 182 administrators responding, 90 (49%) indicated their schools had no policy in place requiring attainment of a specific E^2 benchmark score prior to graduation. These schools indicated the E^2 was not used as part of a progression policy, but functioned as a diagnostic tool to guide student preparation for the NCLEX-RN.

The remaining 92 (51%) RN (BS, AD, and diploma) programs reported having a policy where at least one, or a combination of several consequences were associated with failure to achieve a specified E^2 benchmark score:

- Denied graduation only: 14 programs (15.2%).
- Required remediation only: 19 programs (20.7 %).
- Weighted portion of the capstone course final grade only: 9 programs (9.7 %).
- Denied graduation until completing required remediation: 38 programs (41.3 %).
- Weighted portion of capstone grade and denied graduation: 2 programs (2.2%).
- Weighted portion of capstone grade and required remediation: 2 programs (2%).

- Weighted portion of capstone grade, denial of graduation until completion of required remediation: 8 programs (8.7%).

An independent samples *t*-test was conducted to evaluate differences in student scores between students who took the E^2 without consequences and students who took the E^2 with consequences. The condition of no consequence occurred when the E^2 score functioned only as a diagnostic tool to guide student preparation for the NCLEX-RN. The condition of consequence was when there was progression policy indicating E^2 scores were either used as a portion of the course grade, as a requirement for graduation or permission to take NCLEX, or for required remediation. The dependent variable was the E^2 score where higher scores represent greater probability for success on NCLEX. Table 4 shows sample means, standard deviations, number of subjects and variances for both groups involved in the study.

Table 4

E^2 Score Demographics: Consequence versus No Consequence Condition

Consequence(s) Present	Number of Schools	Number of Subjects	E^2 score		
			M	*SD*	Variance
No	90 (49%)	5,355 (55%)	83.32	11.23	126.12
Yes	92 (51%)	4,340 (45%)	85.89	10.41	108.46

The independent sample *t*-test assumptions of normality and homogeneity of variance (HOV), were evaluated using evaluated using histograms, box plots, the Hartley's F_{max} statistic (Maxwell & Delaney, 2004). [The Levene statistic commonly used by researchers to evaluate homogeneity of variance tends to be sensitive to large sample

sizes and violations of normality (Maxwell & Delaney, 2004). Consequently the Levene reports false positive / significant findings (indicating lack of homogeneity of variance) between groups such as found in this research study containing thousands of subjects (Maxwell & Delaney, 2004). For this reason. Hartley's F_{max} statistic was used to demonstrate homogeneity of variance for this study. The F_{max} is calculated as the ratio of largest within group variance to smallest within group variance (Maxwell & Delaney, 2004). Homogeneity of variance can be assumed when the calculated ratio is a value less than three (Maxwell & Delaney, 2004).

The histograms demonstrated a positive skew with a high frequency of E^2 scores at 99.99. These findings are to be expected since the E^2 exam is predictive of NCLEX outcome and is administered in the last semester of the nursing program. The *t*–test is robust to violations of normality (Green & Salkind, 2004). The box plots showed an equal variance between the two conditions as did the Hartley's F_{max} statistic of 1.16. The assumption of independence was met in that each student score was an individual score from a first time testers. The confidence level for this test was set at .05.

The *t*–test was significant, [t (9693) = -11.60; p < . 0005] indicating that students perform better on the E^2 when consequences are attached than when there are no consequences. The 95% confidence interval for the difference in means was -3.01 to -2.14.

Student Performance When E^2 Scores Required for Graduation

The fourth research question addressed whether nursing students perform differently on the E^2 among programs that specify one of three ranges of a designated benchmark E^2 score as requirement for graduation described as 90, 85, or 80 and below.

Program administrators who indicated their school's progression policy was used for progression to graduation and / or permission to take NCLEX were asked to indicate what E^2 score was used as their program's required benchmark by selecting one of the following: (1) 90; (2) 85; (3) 80; (4) 75; (5) 70; or (6) other, please specify. Of the 182 participating program administrators, fifty-two (28.6%) indicated having a progression policy requiring a certain benchmark score for progression to graduation or eligibility for NCLEX candidacy. Twelve (23.1%) of the 52 schools required an E^2 score of 90; thirty-four (65.4%) schools required an E^2 score of 85; four (7.7%) schools required an E^2 score of 80; one (1.9%) school required an E^2 score of 75; and one (1.9%) school required an E^2 score of 70. One school responded to the "other" category stating that an E^2 score of 90 was required for first time testing, and then E^2 score 85 was required for repeat testing. This school was categorized as one of the twelve schools in the 90 category because only first time test scores were used in this study. The programs who selected an E^2 score of 80, 75 and 70 were combined to form the "80 and below" category, yielding a total of six (11.5%) schools for this category.

A one-way independent samples analysis of variance (ANOVA) was used to evaluate differences between the three conditions of E^2 scores (90, 85, or 80 and below) of designated benchmark E^2 scores used as a requirement for graduation or permission to take NCLEX. The dependent variable was E^2 score where higher scores represent greater probability for success on NCLEX. Table 5 shows estimated populations means, standard deviations, number of subjects and variances for the three conditions of E^2 score.

Table 5

Demographics of Benchmark E^2 Scores Required for Graduation or NCLEX.

E^2 Benchmark	Number of Schools	Number of Subjects	E^2 score		
			M	*SD*	Variance
90	12 (23%)	407 (16%)	85.69	9.78	95.63
85	34 (65%)	1,686 (68%)	86.17	10.32	106.48
80 and below	6 (12%)	401 (16%)	88.33	8.89	79.05

The ANOVA assumptions of normality and homogeneity of variance (HOV), were evaluated using histograms, box plots, and the Hartley's F_{max} statistic. The histograms demonstrated a positive skew with a high frequency of E^2 scores at 99.99. These findings are to be expected since the E^2 exam is predictive of NCLEX outcome and is administered in the last semester of the nursing program. The ANOVA is robust to violations of normality (Maxwell & Delaney, 2004). The box plots showed an equal variance between the 90 condition and 80 and below condition, with a slightly wider variance for the 85 condition. However, the Hartley's F_{max} statistic of 1.35 indicated that the variances were similar enough to assume homogeneity of variance (Maxwell & Delaney, 2004). The assumption of independence was met in that each student score was an individual score representing first time testing. The confidence level for this test was set at .0125.

The ANOVA was significant, [F (2, 2491) = 8.84; $p < .0005$]. The strength of relationship between condition of required benchmark for graduation or permission to

take NCLEX and E^2 score assessed by η^2, was small, with the benchmark condition accounting for 0.7% of the variance in E^2 scores.

Follow-up tests were conducted to evaluate pairwise differences among the means. Because the variances among the three groups were similar, the post-hoc tests were conducted using the Tukey at a .0125 confidence level. There was a significant difference in the means between the 90 benchmark group and the 80 and below benchmark group [t (806) = -3.74; p = .001]; and a significant difference in the means between the 85 benchmark group and the 80 and below benchmark group [t (2,085) = -3.87; p < .0005]. However, there was no difference in student E^2 performance found between the 90 benchmark group and the 85 benchmark group [t (2,091) = -.871; p = .658]. Students at schools having a benchmark set at 80 and below had higher E^2 scores than students at schools having either an 85 or 90 benchmark.

Student Performance When E^2 Scores Calculate Portion of Course Grade

The fifth research question addressed whether nursing students perform differently on the E^2 when programs specify one of four ranges of capstone course final grade weights for the E^2 score described as 1-5%, 6-10%, 11-15%, and 16-20%. Of the 182 program administrators who administered the E^2 during the 2001 -2002 academic year, 23 (12.6%) indicated using the E^2 score as a weighted portion of the capstone course final grade. Administrators were asked to select a range indicating what percentage of the grade was weighted with the E^2 score: (1) 1-5%; (2) 6-10%; (3) 11-15%; (4) 16-20%; or (5) other, please specify. Of the twenty-three programs that indicated using the E^2 score as a weighted portion of the capstone course final grade, sixteen (70%) responded to this question. Seven (43.8%, 1 BSN and 6 ADN) programs

indicated that the E^2 score was weighted between 1-5% of the total course grade; four (25%; 1 BSN and 3 ADN) programs indicated 6-10%; no programs selected the 11-15% category, and four (25%; 1 BSN and 3 ADN) programs weighted the E^2 as 16-20% of the course grade. One (6.3%) program selected "other" for this question, stating the E^2 "replaced the final exam", without providing a percentage weight; therefore, this response was excluded from the analysis.

A one-way independent samples ANOVA was used to evaluate differences between the three conditions of E^2 scores (1-5%, 6-10%, and 16-20%) of designated benchmark E^2 scores used as a weighted portion of the course grade. [Note: Only three ranges (1-5%, 6-10%, and 16-20% were used for the statistical analysis since the other two categories (11-15% and "other") had no subjects. These two categories were dropped from the analysis]. The dependent variable was E^2 score where higher scores represented greater probability for success on NCLEX. Table 6 shows estimated populations means, standard deviations, number of subjects and variances for the three conditions of E^2 score.

Table 6

Demographics of Benchmark E^2 Scores Used as Weighted Portion of Course Grade.

Weighted Portion of Course Grade	Number of Schools	Number of Subjects	E^2 score		
			M	*SD*	Variance
1-5%	7 (47%)	468 (56.6%)	85.29	10.17	103.36
6-10%	4 (27%)	181 (21.9%)	87.44	8.9	79.28
16-20%	4 (27%)	178 (21.5%)	85.14	9.98	99.65

The ANOVA assumptions of normality and homogeneity of variance (HOV), were evaluated using histograms, box plots, and the Hartley's F_{max} statistic. Histograms were positively skewed with an abnormally high frequency of E^2 scores at 99.99. These findings are to be expected since the E^2 exam is predictive of NCLEX outcome and is administered in the last semester of the nursing program. The ANOVA is robust to violations of normality (Maxwell & Delaney, 2004). The box plots showed the spread of each course grade weight range, though not identical, visually looked closely enough related for homogeneity of variance to be tenable. The Hartley's F_{max} statistic of 1.30 indicated that the variances were similar enough to assume homogeneity of variance.

The analysis of variance for global significance was not significant [F (2, 824) = 3.5; p = .031] at the .0125 confidence level. These findings show that there are no significant differences in performance among students whose E^2 scores are used to calculate their final course grade within any of the three ranges of 1 – 5%, 6 – 10%, or 16 – 20%. Despite differences in how much the E^2 score counted towards percent calculation of the course grade, students scores among the three groups were essentially the same.

Student Performance When E^2 Scores Required for Remediation

The sixth research question addressed whether nursing students performed differently on the E^2 when programs specified one of three ranges of a designated benchmark E^2 score for required remediation described as 90, 85, or 80 and below. Program administrators who indicated their school's progression policy was used as a requirement for remediation were asked to indicate what E^2 score was used as their program's required benchmark by selecting one of the following: (1) 90, (2) 85, (3) 80,

(4) 75, (5) 70, (6) other, please specify. Of the 182 participating program administrators, sixty-eight (37.4%) indicated having a progression policy requiring a certain benchmark score for progression to graduation or eligibility for NCLEX candidacy. Ten (15%) of the 68 schools required an E^2 score of 90; forty-nine (72%) schools required an E^2 score of 85; three (4.4%) schools required an E^2 score of 80; five (7.4%) schools required an E^2 score of 75; and one (1.4%) school required an E^2 score of 70. No schools responded to the "other" category. The programs who selected an E^2 score of 80, 75 and 70 were combined to form the "80 and below" category, yielding a total of nine (13.2%) schools for this category.

A one-way independent samples analysis of variance (ANOVA) was used to evaluate differences between the three conditions of E^2 scores (90, 85, or 80 and below) of designated benchmark E^2 scores used as a requirement for remediation. The dependent variable was E^2 score where higher scores represent greater probability for success on NCLEX. Table 7 shows estimated populations means, standard deviations, number of subjects and variances for the three conditions of E^2 score.

Table 7

Demographics of Benchmark E^2 Scores Required for Remediation.

E^2 Benchmark	Number of Schools	Number of Subjects	E^2 score		
			M	*SD*	Variance
90	10 (15%)	314 (10%)	88.28	9.16	83.91
85	49 (72%)	2264 (75%)	85.93	10.5	110.32
80 and below	9 (13%)	454 (15%)	84.45	10.46	109.35

The ANOVA assumptions of normality and homogeneity of variance (HOV), were evaluated using histograms, box plots, and the Hartley's F_{max} statistic. The histograms demonstrated a positive skew with a high frequency of E^2 scores at 99.99. These findings are to be expected since the E^2 exam is predictive of NCLEX outcome and is administered in the last semester of the nursing program. The ANOVA is robust to violations of normality (Maxwell & Delaney, 2004). The box plots showed an equal variance between the 85 condition and 80 and below condition, with a narrower variance for the 90 condition. However, the Hartley's F_{max} statistic of 1.31 indicated that the variances were similar enough to assume homogeneity of variance (Maxwell & Delaney, 2004). The assumption of independence was met in that each student score was an individual score representing first time testing. The confidence level for this test was set at .0125.

The ANOVA was significant, [F (2, 3029) = 12.73; $p < .0005$]. The strength of relationship between condition of required benchmark for graduation or permission to take NCLEX and E^2 score assessed by η^2, was small, with the benchmark condition accounting for 0.8% of the variance in E^2 scores.

Follow-up tests were conducted to evaluate pairwise differences among the means. Because the variances among the three groups were similar, post-hoc tests were conducted using the Tukey at a .0125 confidence level. There was a significant difference in the means between the 90 benchmark group and the 80 and below benchmark group [t (767) = 1.31; $p <.0005$]; and a significant difference in the means between the 90 benchmark group and the 85 benchmark group [t (2577) = 3.77; p = .001]. However, there was no difference in student E^2 performance found between the

85 benchmark group and the 80 and below benchmark group [t (2717) = 2.8; p = .015]. In terms of benchmarking for mandatory remediation, students required to achieve a benchmark of 90 scored higher on the E^2 than students required to achieve the 85 or the 80 and below benchmarks.

Student E^2 Performance with One, Two or Three Consequences

The seventh research question addressed whether nursing students perform differently on the E^2 in the presence of one, two or three consequences. Ninety-two (51%) of the 182 program administrators indicated having some type of progression policy that tied E^2 score to a consequence or combination of consequences. Based on responses, programs were categorized into one of three groups: one consequence, two consequences or three consequences. The one consequence group was comprised of 2,028 (47%) student scores from 42 (46%) schools who specified the E^2 score for either denial of graduation or permission to take NCLEX only (14 programs); requirement for remediation only (19 programs) or weighted portion of the capstone course final grade only (9 programs). The two consequence group was comprised of 1,785 (41%) student scores from 42 (46%) schools who specified a combination of two consequences associated with E^2 score; either as a benchmark for denial of graduation until completion of required remediation (38 programs), as a weighted portion of capstone grade and as a benchmark for denial of graduation (2 programs), or as a weighted portion of capstone grade and as a benchmark for required remediation (2 programs). The three consequence group comprised 527 (12%) student scores from 8 (8%) schools who specified the E^2 score for all three types of consequences including weighted portion of

the capstone course final grade, denial of graduation until completion of required remediation.

A one-way independent samples analysis of variance (ANOVA) was used to evaluate differences between the three conditions of combination of consequences (one, two or three) associated with E^2 scores. The dependent variable was E^2 score where higher scores represent greater probability for success on NCLEX. Table 8 shows estimated populations means, standard deviations, number of subjects and variances for the three conditions of E^2 score.

Table 8

Demographics of E^2 Scores Associated With One, Two, or Three Consequences.

Number of Consequences	Number of Schools	Number of Subjects	E^2 score *M*	*SD*	Variance
One	42 (46%)	2028 (47%)	84.72	10.84	117.45
Two	42 (46%)	1785 (41%)	87.35	9.78	95.69
Three	8 (8%)	527 (12%)	85.54	10.24	104.87

The ANOVA assumptions of normality and homogeneity of variance (HOV), were evaluated using histograms, box plots, and the Hartley's F_{max} statistic. The histograms demonstrated a positive skew with a high frequency of E^2 scores at 99.99. These findings are to be expected since the E^2 exam is predictive of NCLEX outcome and is administered in the last semester of the nursing program. The ANOVA is robust to violations of normality (Maxwell & Delaney, 2004). The box plots showed equal variance between all three conditions and the Hartley's F_{max} statistic of 1.23 indicated

that the variances were similar enough to assume homogeneity of variance (Maxwell & Delaney, 2004). The assumption of independence was met in that each student score was an individual score representing first time testing. The confidence level for this test was set at .0125.

The ANOVA was significant, [F (2, 4337) = 31.10; p < .0005]. The strength of relationship between condition of to take NCLEX and E^2 score assessed by η^2, was small, with the benchmark condition accounting for 1.4% of the variance in E^2 scores.

Follow-up tests were conducted to evaluate pairwise differences among the means. Because the variances among the three groups were similar, post-hoc tests were conducted using the Tukey at a .0125 confidence level. There was a significant difference in the means between the one consequence group and the two consequence group [t (3812) = 7.84; p <.0005]; and a significant difference in the means between the two consequence group and the three consequence group [t (2311) = 3.53; p = .001]. However, there was no difference in student E^2 performance found between the one consequence group and the three consequence group [t (2554) = 1.62; p = .24]. Students having two or more consequences made higher E^2 scores than students having only one consequence.

Summary of the Findings

The E^2, using the weighted score calculation method to predict NCLEX outcome, demonstrated a high degree of predictive accuracy in predicting NCLEX success (97.8%) for students who took the exam during the 2001-2002 academic year. The predictive accuracy of the E^2 for BS student performance on NCLEX was 97%. For AD students it was 98.4%, and for diploma students it was 96.2%. 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. No significant difference was found between AD and BS programs when they were compared at like intervals.

Regarding whether nursing students performed differently on the E^2 in the presence of consequences as mandated by progression policies, analysis indicated significantly better student performance on the E^2 when consequences were attached compared to when there were no consequences. Nursing student performance was significantly different among programs that specified one of three ranges of a designated benchmark E^2 score as a requirement for graduation described as 90, 85 and 80 and below. Specifically, the analysis showed that students in the 80 and below group performed significantly better than students in both the 90 group and the 85 group; but there was no difference in performance between students in the 90 group or 85 group.

In relation to student E^2 performance when scores were used to calculate a portion of the nursing student's capstone course final grade, there were no significant differences in performance among students within any of the three ranges of 1-5%, 6-10%, or 16-20%. However, there were significant differences among student performance on the E^2 when scores were benchmarked at either 90, 85, or 80 and below, and used to require remediation prior to graduation. Students required to achieve an E^2 score of 90 performed significantly better than students required to achieve an 85 or students required to achieve an 80 and below; but there were no differences in E^2

scores between students with the 85 E^2 score requirement and students with the 80 and below requirement.

Students performed differently on the E^2 based on the number of consequences (1, 2, or 3) associated with their E^2 score. Students with 2 consequences associated with their score performed significantly better than students with only one consequence or students with three consequences. There were no differences in performance between students in the one consequence and the three consequence groups.

CHAPTER 5

SUMMARY OF THE STUDY

The purposes of this study were to: 1) assess the accuracy of Health Education Systems Inc. (HESI)'s Exit Exam (E^2) in predicting National Council of State Boards of Nursing Examination for Registered Nurses (NCLEX-RN) success based on the HESI Predictability Model's (HPM) new individual weighted item difficulty scoring method; 2) identify differences in NCLEX outcomes for students whose E^2 scores are in the A/B, C, D, E/F, and G/H scoring intervals; 3) identify differences between mean E^2 scores of students enrolled in nursing programs that have adopted benchmark progression policies and those enrolled in programs that have not adopted such policies; 4) identify differences between mean E^2 scores among programs that specify one of three ranges of a designated benchmark E^2 score as requirement for graduation described as 90, 85 or 80 and below; 5) identify differences between mean E^2 scores among programs that specify one of four ranges of capstone course final grade weights for the E^2 scores described as 1-5%, 6-10%, 11-15%, and 16-20%; 6) identify differences in mean E^2 scores among programs that specify one of three ranges of a designated benchmark E^2 score for required remediation described as 90, 85, or 80 and below; and 7) identify differences in mean E^2 scores among students in nursing programs that use a single benchmark consequence or a combination of two or three benchmarking consequences. Components of two theoretical frameworks helped guide this study: classical test theory and Weiner's (1986) attributional theory of motivation and emotion. Classical test theory was the basis for determining the E^2's predictive accuracy; and Weiner's attributional

theory of motivation and emotion established the framework to study the impact of progression policies on student performance. This chapter contains a summary of the investigation, a discussion of findings, investigational conclusions, implications, and recommendations for further study.

Summary

A descriptive, retrospective design was used to examine the data provided by schools of nursing regarding students' NCLEX outcomes and progression policy practices in the academic year 2001-2002. A total of 14,727 students took the E^2 during Year V. Data were obtained from administrators or their designees at RN schools of nursing that administered the E^2 in year V and responded to the questionnaire used to build the HESI database. HESI staff obtained information for the database by sending schools that administered the E^2, an electronic mail message inviting participation in the study. School administrators clicked on an electronic link in the message and were given a summary analysis of their aggregate data grouping students' scores by scoring categories ranging from A/B, the highest-scoring category, to G/H, the lowest scoring category. Administrators were asked to identify from the list, the number of students who failed NCLEX. A second portion of the survey questioned respondents regarding their school's use of the HESI E^2 in relation to a progression policy, asking for identification of specific uses of student E^2 scores in their nursing program. Responses to the on-line survey were immediately saved in the HESI database. HESI removed school and student names from the data to preserve confidentiality and then provided the researcher with aggregate student scores by school, NCLEX pass rates and program progression policy practices.

Of the 277 program administrators invited to participate in the study, 182 (65.7%) submitted responses describing administrations of the E^2 to 9,695 students. Of the 182 administrators who provided HESI with student NCLEX outcome and progression policy information, 90 (49%) indicated their schools had no policy in place requiring attainment of a specific E^2 benchmark score prior to graduation. These schools indicated the E^2 was not used as part of a progression policy, but functioned as a diagnostic tool to guide student preparation for the NCLEX-RN.

The remaining 92 (51%) RN programs reported having a policy where at least one, or a combination of several consequences were associated with failure to achieve a specified E^2 benchmark score:

- Denied graduation only; 14 programs (15.2%).
- Required remediation only; 19 programs (20.7 %).
- Weighted portion of the capstone course final grade only; 9 programs (9.7 %).
- Denied graduation until completing required remediation; 38 programs (41.3 %).
- Weighted portion of capstone grade and denied graduation; 2 programs (2.2%).
- Weighted portion of capstone grade and required remediation; 2 programs (2.2%).
- Weighted portion of capstone grade, denial of graduation until completion of required remediation; 8 programs (8.7%).

The overall predictive accuracy, research question 1, was 97.8% for the 2001 – 2002 academic year. The predictive accuracy of the E^2 for BS student performance on NCLEX was 97%. For AD students it was 98.4%, and for diploma students it was 96.2%.

Analysis for differences between scoring intervals, research question 2, indicated that NCLEX-RN failures increased as the scoring interval decreased. In addition, analysis for differences between mean E^2 scores of students enrolled in nursing programs that have adopted benchmark policies and those enrolled in programs that have not adopted such policies, research question 3, indicated that students perform better on the E^2 when consequences are attached than when there are no consequences.

Student performance was also found to vary significantly based on severity of the different types consequences including requirement for graduation, weighted portion of the course grade, and requirement for remediation as addressed by research questions 4, 5 and 6 respectively. When the E^2 score was required for graduation, (research question 4) students at schools having a benchmark set at 80 and below, had higher E^2 scores than students at schools having either an 85 or 90 benchmark. When the E^2 score was used to calculate portion of the course grade, (research question 5) students' scores among the three groups (1-5%, 6-10%, and 16-20%) were essentially the same. When the E^2 score was required for remediation, (research question 6) students required to achieve a benchmark of 90 scored higher on the E^2 than students required to achieve the 85 or the 80 and below benchmarks.

Finally, analysis for research question 7 evaluated differences in student performance based on the number of consequences (1, 2, or 3) associated with their E^2 score. Students having two or more consequences made higher E^2 scores than students having only one consequence.

Discussion of the Findings

Predictive Accuracy of the E^2 Regarding NCLEX Success

The E^2 demonstrated strong predictive ability of NCLEX success (97.9%) for the fifth consecutive year. The predictive accuracy of year V was similar to the findings of the four previous years, indicating that based on aggregate data collected from 27,563 subjects for five years, of all the standardized tests administered to students during the later half of the student's senior year, the E^2 continues to be the best predictor of NCLEX success (Lauchner, Newman, & Britt, 1999; Newman, Britt, & Lauchner, 2000; Nibert, 2003; Nibert & Young, 2001). Prediction of NCLEX success has been and continues to be important for nursing programs since accreditation, enrollment, funding and reputation are based on NCLEX pass rates. NCLEX success is also important for nursing graduates taking the exam, since those who fail are unable to obtain employment, and may experience a sense of low self-esteem (Billings et al., 1996). Of the 42 research studies (see Table 1) identifying 89 NCLEX predictor variables from 1990 – 2005, the E^2 has clearly proven its ability to predict NCLEX success. The E^2 is the most effective tool for identifying at risk students, allowing faculty the chance to implement remediation strategies prior to NCLEX testing. Developers of the E^2 and users alike claim the exam's main purpose is to assess student's readiness for the NCLEX (Morrison, Adamson, Nibert, & Hsia, 2004), pinpointing student's subject content weaknesses to provide an invaluable asset in designing individualized remediation programs (Engelmann & Caputi, 1999; Morrison, Free, & Newman, 2002).

In schools that use the E^2 for remediation, significantly fewer low-scoring students have been found to fail the licensure exam with faculty attributing this finding to student

participation in remediation (Newman, Britt, & Lauchner, 2000). The Exit Exam's strong predictive ability provides faculty assurance in implementing progression policies based on E^2 score to improve and / or strengthen NCLEX outcomes.

Differences in NCLEX Outcomes by HESI Scoring Intervals

Analysis of students' E^2 performance by specific scoring intervals yielded a consistent stair-step pattern where the percentage of students who failed the NCLEX significantly increased with each decent to a lower scoring interval. This finding is consistent with Nibert, Young and Britt's (2003) analysis of student E^2 performance during the 1999-2000 academic year. Nibert, et al. also compared NCLEX outcomes for the five HESI scoring intervals, and found that NCLEX failures increased as the scoring interval decreased (Nibert et al., 2003). This pattern helps faculty establish E^2 score benchmarks, design progression policies and implement remediation strategies to address nursing student learning and NCLEX preparation needs according to their indicated risk.

Comparisons between BS and AD students within each E^2 scoring interval category yielded no differences in NCLEX-RN outcomes, indicating that the E^2 is an effective predictor of NCLEX success for both program types.

Student E^2 Performance with and Without Progression Policy Consequences

Students performed significantly ($p < .0005$) better on the E^2 in the presence of one or more consequences (denial of graduation, required remediation and weighted portion of the capstone grade) than when taking the E^2 with no consequence tied to the score. Progression policies requiring achievement of a designated benchmark score may have caused nursing students to perceive the E^2 as a high stakes exam similar to

NCLEX. These findings support Weiner's (1986) attributional theory of motivation and emotion which proposes that achievement or affiliation needs influence perceptions, perceptions generate achievement needs, and expectancy of success is associated with performance intensity of achievement strivings. According to Weiner's theory (1986), significantly higher E^2 scores demonstrate that students perceiving the E^2 to be a high stakes exam because of benchmark consequences, experience an increased expectancy of success and have greater performance strivings to succeed on the exam than students who have no benchmark consequences.

Faculty administering the E^2 without significant consequences attached may be concerned that students are not taking the exam seriously, or are not trying their best on the exam. Of the 182 schools represented in this study, 90 (49%) indicated no consequence was tied to student performance on the E^2. Findings from this study support previously cited literature showing variance in student performance on standardized examinations associated with student motivation levels (Burke, 1991; Schmidt & Ryan, 1992; Smith & Smith, 2002; Sundre, 1997; Sundre & Kitsantas, 2004; Wolf & Smith, 1995; Wolf, Smith, & Birnbaum, 1995); and performance variance occurring in the absence of consequences for examinees (Sundre, 1999). Nursing faculty administering the E^2 without associated consequences may want to consider how the disposition of test-takers impacts performance. Students who approach the E^2 as "just another exam" or perceive the test results as having no significant bearing, may not feel motivated to perform well, despite having sufficient knowledge (Pintrich & DeGroot, 1990). Faculties may want to consider increasing student motivation for optimal E^2

performance by establishing a progression policy associating one or more consequences with student E^2 performance.

Faculties deciding to establish a progression policy may want to consider previously reported literature stating that implementing a progression policy tailored to the individual needs of a program's nursing student population may prove somewhat of a challenge (Morrison et al., 2002; Nibert, 2003). When designing progression policies, faculty are encouraged to consider current pass rates, program size, type and number of consequences, and level of designated benchmark, if any, that would be best support their program objectives (Nibert, 2003).

Student Performance When E^2 Scores Required for Graduation

Infrequently used lower benchmarks (80 and below), [used by only 6 (12%) of the 52 schools], resulted in better student performance on the E^2 than either the 85 or the 90 benchmarks. An explanation for this outcome is that policies stipulating a benchmark score of 85 or 90 may have been established by programs with historically lower NCLEX-RN pass rates. These programs may have set higher benchmark requirements in order to improve NCLEX-RN outcomes of their graduates. For small programs, or programs with declining pass rates, even the NCLEX-RN failure of a single student could prove devastating to the program's annual NCLEX-RN pass rate. Therefore, stipulating a higher benchmark may have been necessary to identify students at-risk of NCLEX-RN failure, and allowed faculties to prevent students in need of remediation from taking the NCLEX-RN until they demonstrated readiness for licensure candidacy by achieving the program's benchmark score on another version of the E^2.

The phenomenon might further explain why mean E^2 scores of students in the 85 and 90 groups did not increase as the benchmark requirement increased. Previous studies indicated that significantly more low-scoring students who took the E^2 failed the NCLEX-RN when compared to high-scoring students (Newman et al., 2000; Nibert & Young, 2001) and that 24% of students scoring in the lowest E^2 scoring interval (70 – 79%) failed the NCLEX-RN (Nibert, Young, & Adamson, 2002). Faculty setting the benchmark at 85 or at 90 potentially recognized that a pattern of an entrenched population of low-scoring students on the E^2 existing within their programs for a protracted period of time placed the programs at risk of repeated below-acceptable NCLEX-RN pass rate levels persisting over many successive years. The 85 or 90 benchmark requirement may have been an attempt to compel the issue of mandatory remediation as a consequence of failure to achieve the benchmark so that more NCLEX-RN failures could be averted within a struggling program.

It is interesting to note that for this research, student E^2 performance was the same regardless of whether faculty designated a benchmark score of 85 or 90. This finding suggests that the level of benchmark designated for progression to graduation or permission to take NCLEX should not be based on student motivation, but on level of risk of NCLEX failure faculty find acceptable to meet program objectives. Findings also suggest that benchmark requirements for progression are not driving E^2 scores, but that E^2 scores are likely driving the trend in increasing adoption of benchmark requirements as a function of progression policies.

Student Performance When E^2 Scores Calculate Portion of Course Grade

No significant differences in performance were found in this research among students whose E^2 scores were used to calculate a portion of their final course grade within any of the three ranges of 1-5%, 6-10%, or 16-20%. This finding suggests that although this type of consequence (weighted portion of the capstone course final grade) associated with the E^2 score may motivate student performance on the E^2, the actual percent of the course grade designated by the progression policy may not make much difference in student performance.

Student Performance When E^2 Scores Required for Remediation

Significant differences ($p < .0005$) were found between the three conditions of designated benchmark E^2 scores (90, 85 and 80 and below) used as a requirement for remediation. The 90 benchmark group for required remediation performed significantly better than both the 85 group ($p = .001$) and the 80 and below group ($p < .005$), but for this research, no difference in E^2 performance was found between the 85 group and the 80 and below groups. These findings did support Weiner's (1986) attributional theory of motivation and emotion suggesting increased difficulty of outcome produces sufficient motivation and achievement strivings as manifested in student E^2 performance. This finding provides guidance for faculties considering using E^2 score as a requirement for remediation. Based on these results, faculties can expect that setting the benchmark score as high as 90 in requiring remediation will likely improve student performance on the E^2 especially when requiring remediation is the only consequence tied to E^2 score. Progression policies only requiring remediation prior to NCLEX will likely prove to facilitate a greater number of students at risk benefiting from remediation strategies

when the benchmark is set as high as 90. Future research is recommended to determine methods of remediation most effective in facilitating maximum improvement of E^2 repeat performance.

Student Performance with One, Two or Three Consequences

Students performed differently ($p < .0005$) on the E^2 in the presence of one, two or three consequences. Specifically, students in the two consequences group performed better than students in the one consequence ($p < .005$) and three consequences ($p = .001$) groups. Of the 52 schools mandating consequences, 42 (46%) indicated using two consequences compared to 42 (46%) who indicated only one consequence tied to E^2 performance and eight (8%) who indicated using three consequences. With an equal number of schools using one and two consequences, it is recommended that faculties consider implementing progression policies mandating two consequences for maximum student motivation and performance on the E^2. The most common combination of two consequences associated with E^2 score used by 38 (90%) of the 42 schools was a benchmark for denial of graduation until completion of required remediation.

Conclusions and Implications

Findings from this investigation support the following conclusions:

1. The E^2 is highly accurate in predicting NCLEX success in registered nurse programs.
2. The percentage of students who failed the NCLEX rises with each successive drop in scoring interval.

3. When consequences are attached to E^2 score students perform better than they would without consequences.
4. Student E^2 performance associated with a progression policy tied to graduation is better for lower benchmarks, than for higher benchmarks.
5. Although consequences, including course grade weight associated with E^2 score, influence student performance, the actual level of course grade percentage weighting does not influence student performance.
6. Students are motivated to perform better on the E^2 when their score is tied to a high benchmark rather than a lower benchmark for required remediation.
7. Progression policies tying two consequences to E^2 score (denial of graduation until completion of required remediation) are associated with the better student performance on E^2 when compared to progression policies delineating only one or as many as three consequences.

Several implications can be derived from this study. Findings regarding the predictive accuracy and differences in NCLEX for the various scoring intervals should provide sufficient evidence for nursing faculty in the decision to utilize the E^2 as a diagnostic tool in determining student preparedness for NCLEX and for identifying student remediation needs prior to graduation.

This research establishes an increase in student motivation in the presence of a progression policy tying E^2 score to consequences such as progression to graduation, a weighted portion of the capstone course final grade and required remediation. Faculties considering implementing any type of progression policy should consider findings regarding the designated level of the benchmark and type and number of consequences

used. While consequences make a difference in student performance on the E^2, schools must make decisions about specific consequences based on the characteristics and needs of their program.

Delineation of the E^2 score as a graduation requirement should be used with caution. Students in the lower benchmark group (80 and below) performed better on the E^2 when a designated score was required for graduation. Higher benchmark requirements for graduation (85 and 90) may cause students excessive anxiety impacting their effectiveness during preparation and performance during administration. Faculty may justify setting higher benchmarks (85 and 90) based on the E^2's highly predictive nature; determining that higher benchmarks more likely prevent greater numbers of at-risk students from failing NCLEX, thus helping maintain program high NCLEX pass rates.

Maintenance of high NCLEX pass rates should not be done solely via benchmarking of E^2 scores. Use of the E^2 score as a single evaluative measure in the last semester may prove devastating to students who have progressed successfully in the program, have spent time, money and effort, and who are unaware of weaknesses until their E^2 score fails to meet graduation requirements. Faculty could also impact / maintain NCLEX pass rates via progressive evaluative mechanisms throughout the nursing program to identify and inhibit at-risk students during the program. Additional evaluative measures would help redirect or eliminate the poor performing, at-risk students prior to taking the E^2 in their last semester. Also, in conjunction with progression policies, faculty should consider other alternatives such as curriculum revisions, augmented testing regimes, use of computerized testing, and additional

learning opportunities promoting critical thinking and problem solving.

The E^2 administered in the final semester would be better used as a last measure of student readiness, geared more as a confirmation of student readiness for NCLEX, and as an indication of potentially troubled areas the student may want to focus on prior to taking NCLEX. The exam is designed to prepare students for NCLEX-type questions and give students experience with the computerized testing format.

Findings in this study fail to provide support in determining what portion of the capstone course final grade the E^2 should contribute since students did not perform differently among the three ranges of percentages used. Despite no differences found in this research, faculty may want to consider using this type of consequence for performance motivation and should recognize that regardless of course grade weight used, the greater percentage contributed toward the final course grade, the greater the impact this exam's outcome will have on the final grade. In other words, student poor E^2 performance contributing toward a significant portion of the course grade, would possibly result in course failure and ineligibility for graduation or NCLEX.

Recommendations for Further Study

Suggestions for future research are as follows:

1. Although the predictive accuracy of the E^2 has been well established over five years of study, periodic evaluations of the exam's ability to predict NCLEX success should be conducted since NCLEX is constantly adapting to reflect changes in health care and competence of new nurses entering the profession. Also calculation of the Exit Exam's ability to predict NCLEX success for international nurses applying for United States licensure should be investigated.

2. Examine the effect of consequences associated with E^2 benchmark achievement on NCLEX outcomes.
3. Compare E^2 benchmarking standards adopted by nursing programs to explore the pattern of both E^2 failure and NCLEX-RN failure in schools requiring achievement of an 85 or 90 E^2 benchmark score.
4. Investigate methodology and reasoning behind how nursing faculties determine required E^2 benchmarks.
5. Examine whether the E^2 is predictive of graduate nurse success in the clinical setting.
6. Investigate effect of progression policies on nursing student responses to high stakes testing.
7. Explore the effectiveness of remediation strategies used in helping students improve E^2 scores.
8. 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.
9. Investigate the predictive accuracy of mid-curricular exams on E^2 performance.

REFERENCES

AACN. (2002). *CCNE accreditation standards*. Retrieved March, 2005, from http://www.aacn.nche.edu/Accreditation/standrds.htm

Alexander, J., & Brophy, G. (1997). A five-year study of graduates' performance on NCLEX-RN. *Journal of Nursing Education, 36*(9), 443 - 445.

Arathuzik, D., & Aber, C. (1998). Factors associated with National Council Licensure Examination--registered nurse success. *Journal of Professional Nursing, 14*(2), 119 - 126.

Arvey, R., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology, 43*, 695-716.

Atkinson, J., & Feather, N. (1966). *A theory of achievement motivation*. New York: Wiley.

Banta, T., & Pike, G. (1989). Methods of comparing outcomes assessment instruments. *Research in Higher Education, 30*(5), 455-469.

Barkley, T., Rhodes, R., & Dufour, C. (1998). Licensure examination. Predictors of success on the NCLEX-RN among baccalaureate nursing students. *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., Hodson-Carlton, K., Kirkpatrick, J., Aaltonen, P., Dillard, N., Richardson, V., et al. (1996). Computerized NCLEX-RN preparation programs: a comparative review. *Computers in Nursing, 14*(5), 272 - 286.

Briscoe, V., & Anema, M. (1999). The relationship of academic variables as predictors of success on the National Council Licensure Examination for Registered Nurses (NCLEX-RN) in a selected associate degree program. *The Association of Black Nursing Faculty, 10*(4), 80 - 83.

Burke, P. (1991). *You can lead adolescents to a test you can't make them try: Final report* (Contract No. OTA-H3-6110.0). Washington, DC: Office of Technology Assessment.

Burns, N., & Grove, S. (1997). *The practice of nursing research: Conduct, critique, & utilization* (3rd ed.). Philadelphia, PA: W. B. Saunders.

Carpenter, D., & Bailey, P. (1999). Predicting success on the registered nurse licensure examination--1999 update. In K. Stevens (Ed.), *Evidence-based teaching: Current research in nursing education.* (pp. 135 - 170). Sudbury, MA: Jones and Bartlett Publishers.

Collins, P. (2002). Predicting a passing outcome on the National Council Licensure Examination for registered nurses by associate degree graduates. *DAI, 64*(01B), 107.

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

Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity.* Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.

Crow, C., Handley, M., Morrison, R., & Shelton, M. (2004). Requirements and interventions used by BSN programs to promote and predict NCLEX-RN success: A national study. *Journal of Professional Nursing, 20*(3), 174-186.

Daley, L., Kirkpatrick, B., Frazier, S., Chung, M., & Moser, D. (2003). Predictors of NCLEX-RN success in a baccalaureate nursing program as a foundation for remediation. *Journal of Nursing Education, 42*(9), 390-398.

Dell, M., & Valine, W. (1990). Explaining differences in NCLEX-RN scores with certain cognitive and non-cognitive factors for new baccalaureate nurse graduates. *Journal of Nursing Education, 29*(4), 158 - 162.

Drake, C., & Michael, W. (1995). Criterion-related validity of selected achievement measures in the prediction of a passing or failing criterion on the National Council Licensure Examination (NCLEX) for nursing students in a two-year associate degree program. *Educational and Psychological Measurement, 55*, 675 - 683.

Endres, D. (1997). A comparison of predictors of success on the NCLEX-RN for African American, foreign-born, and white baccalaureate graduates. *Journal of Nursing Education, 36*(8), 365 - 371.

Engelmann, L., & Caputi, L. (1999). *Coaching your students to NCLEX success.* Glen Ellyn, IL: College of DuPage.

Foti, I., & DeYoung, S. (1991). Predicting success on the National Council Licensure Examination-Registered Nurse: another piece of the puzzle. *Journal of Professional Nursing, 7*(2), 99 - 104.

Fowles, E. (1992). Predictors of success on NCLEX-RN and within the nursing curriculum: Implications for early intervention. *Journal of Nursing Education, 31*(2), 53 - 57.

Freud, S. (1920). *Beyond the pleasure principle* (Vol. 18). London: Hogarth Press.

Frisbe, D. A. (1988). Reliability of scores from teacher-made tests. *Educational Measurement: Issues and Practice, 7*(1), 25-35.

Gabbin, A. (2002). Experience with using samples to assess students in the major. *Assessment Update, 14*(2), 6-7, 9.

Gallagher, P., Bomba, C., & Crane, L. (2001). Using an admissions exam to predict student success in an ADN program. *Nurse Educator, 26*(3), 132 - 135.

Giddens, J. (2002). The relationship of critical thinking to performance on the National Council Licensure Examination for Registered Nurses (NCLEX-RN). *DAI, 63*(12B), 106.

Green, S. B., & Salkind, N. J. (2004). *Using SPSS for Windows and Macintosh analyzing and understanding data.* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Griffiths, M., Papastrat, K., Czekanski, K., & Hagan, K. (2004). The lived experience of NCLEX failure. *Journal of Nursing Education, 43*(7), 322-325.

Haas, R., Nugent, K., & Rule, R. (2004). The use of discriminant function analysis to predict student success on the NCLEX-RN. *Journal of Nursing Education, 43*(10), 440-446.

Hanks, C. (1999). How well does a test predict an outcome? *Computers in Nursing, 17*(6), 241 - 243.

Henriques, M. (2002). Predictors of performance on the National Council Licensure Examination for Registered Nurses. *DAI, 63*(11B), 85.

HESI. (2003). Health Education Systems, Incorporated Database. Houston: Health Education Systems Inc.

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.

Horns, P., O'Sullivan, P., & Goodman, R. (1991). The use of progressive indicators as predictors of NCLEX-RN success and performance of BSN graduates. *Journal of Nursing Education, 30*(1), 9 - 14.

Huck, S. W. (2000). *Reading statistics and research* (3rd ed.). New York: Addison Wesley Longman, Inc.

Hull, C. L. (1943). *Principles of behavior*. New york: Appleton-Century-Crofts.

Jakwerth, P., & Stancavage, F. (2003). *An investigation of why students do not respond to questions NAEP VAlidity Studies. Working Paper Series.* (Reports - Research (143) No. NCES-WP=2003-12). Washington, DC: National Center for Education Statistics.

Julian, E. P., Wendt, A. P. R. N., Way, D. P., & Zara, A. P. (2001). Moving a National Licensure Examination to Computer. [Miscellaneous]. *Nurse Educator November/December, 26*(6), 264-267.

Kehoe, J. (1995). *Basic item analysis for multiple-choice tests.* Retrieved August 5, 2005, from http://PAREonline.net/getvn.asp?v=4&n=10 .

Lamm, G., & McDaniel, A. (2000). Research briefs. Factors that predict success on NCLEX-PN. *Journal of Nursing Education, 30*(1), 9 - 14.

Lauchner, K., Newman, M., & Britt, R. (1999a). How well does a test predict an outcome? *Computers in Nursing, 17*(6), 243-246.

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

Lengacher, C., & Keller, R. (1990). Academic predictors of success on the NCLEX-RN examination for associate degree nursing students...National Council Licensing Examination for Registered Nurses. *Journal of Nursing Education, 29*(4), 163 - 169.

Lewin, K. (1935). *A dynamic theory of personality*. New York: McGraw-Hill.

Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Prentice Hall.

Lipsey, M. (1990). *Design sensitivity statistical power for experiemental research.* Newbury Park, CA: Sage.

Matassarin-Jacobs, E. (1989). The nursing licensure process and the NCLEX-RN. *Nurse Educator, 14*(6), 32 - 35.

Maxwell, S., & Delaney, H. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahway, NJ: Lawrence Erlbaum Associates.

McClelland, E., Yang, J., & Glick, O. (1992). A statewide study of academic variables affecting performance of baccalaureate nursing graduates on licensure examination. *Journal of Professional Nursing, 8*(6), 342 - 350.

McQueen, L., Shelton, P., & Zimmerman, L. (2004). A collective community approach to preparing nursing students for the NCLEX RN examination. *ABNF Journal, 15*(3), 55-59.

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.

Milan, C. L. (1997). Identifying predictors of performance for the associate degree graduate nurse on the National Council Licensure Examination for Registered Nurses. *DAI, 58*(08B), 96.

Mills, A., Becker, A., Sampel, M., & Pohlman, V. (1992). Success-failure on the National Council Licensure Examination for Registered Nurses by nurse candidates from an accelerated baccalaureate nursing program. *Journal of Professional Nursing, 8*(6), 351 - 357.

Mills, A., Sampel, M., Pohlman, V., & Becker, A. (1992). The odds for success on NCLEX-RN by nurse candidates from a four-year baccalaureate nursing program. *Journal of Nursing Education, 31*(9), 403 - 408.

Morrison, S. (2000, July 13, 2000). *Recommendations for improving NCLEX-RN pass rate*. Retrieved May 17, 2005, from http://www.hesitest.com/showNewsletter.asp?imageid=6

Morrison, S., Adamson, C., Nibert, A., & Hsia, S. (2004). HESI exams an overview of reliability and validity. *CIN: Computers, Informatics, Nursing, 22*(4), 220-226.

Morrison, S., & Free, K. (2001). Writing multiple-choice test items that promote and measure critical thinking. *Journal of Nursing Education, 40*(1), 17 - 24.

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., Smith, P., & Britt, R. (1996). *Critical Thinking and Test Item Writing.* Houston, TX: Health Education Systems, Inc.

Napoli, A., & Raymond, L. (2004). How reliable are our assessment data?: A comparison of the reliability data produced in graded and un-graded conditions. *Research in Higher Education, 45*(8), 921-929.

National Council of State Boards of Nursing. (2001). *National Council Detailed Test Plan for the NCLEX-RN Examination.* Chicago: National Council of State Boards of Nursing.

National Council of State Boards of Nursing. (2004). *Testing services NCLEX psychometrics: NCLEX pass rates.* Retrieved May 26, 2005, from http://www.ncsbn.org/testing/psychometrics_nclexpassrates.asp

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. (2003). *Predicting NCLEX success with the HESI Exit Exam: Results from four years of study.* Unpublished Dissertation, Texas Woman's University, Denton, Texas.

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

Nibert, A., Young, A., & Adamson, C. (2002). Predicting NCLEX success with the HESI Exit Exam. *Computer Informatics Nursing, 20*(6), 261 - 267.

Nibert, A., Young, A., & Britt, R. (2003). The HESI Exit Exam: Progression benchmark and remediation guide. *Nurse Educator, 28*(3), 141 - 146.

NLNAC. (2004). *Accreditation manual with interpretive guidelines by program type for postsecondary and higher degree programs in nursing.* Retrieved March, 2005, from http://www.nlnac.org/manuals/Manual2004.htm

Nnedu, C. C. (2000). Academic predictors of success for baccalaureate students on the National Council Registered Nurse Licensure Examination. *DAI, 61*(09B), 135.

Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory*. New York: McGraw-Hill.

Olsen, S., & Wilson, K. (1991). *A follow-up of suspect sophomore scores on the COMP test*. Kirkville, Missouri: Northeast Missouri State University.

Ostrye, M. (2001). Predicting NCLEX-PN performance for practical nursing students. *Nurse Educator, 26*(4), 170 - 174.

Percoco, T. A. (2001). Variables predictive of program and NCLEX success for associate degree nursing students. *DAI, 62*(08B), 78.

Pintrich, P., & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *J Educ Psychol, 82*, 33 - 40.

Poorman, S., & Martin, E. (1991). The role of nonacademic variables in passing the National Council Licensure Examination. *Journal of Professional Nursing, 7*(1), 25 - 32.

Rotter, J. B. (1954). *Social learning and clinical psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Schaal, M. (1990). *An investigation of the predictability of the performance of graduates of a baccalaureate nursing program on the NCLEX-RN using cognitive and noncognitive variables.*, Rutgers, New Brunswick.

Schafer, V. (2002). Determining the predictors of success for nursing students and graduates of MidAmerica Nazarene University. *DAI, 64*(02B), 487.

Schmidt, M., & Ryan, A. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology, 77*(5), 629-637.

Schwirian, P., Baer, C., Basta, S., & Larabee, J. (1978). *Prediction of successful nursing performance* (No. Publication No. 0-245-048). Washington, DC: U. S. Government Printing Office.

Seldomridge, L., & DiBartolo, M. (2004). Can success and failure be predicted for baccalaureate graduates on the computerized NCLEX-RN? *Journal of Professional Nursing, 20*(6), 361-368.

Smith, L. F., & Smith, J. K. (2002). Relation of test-specific motivation and anxiety to test performance. *Psychological Reports, 91*(3,Pt1), 1011-1021.

Sundre, D. L. (1997, April). *Differential examinee motivation and validity: A dangerous combination.* Paper presented at the American Educational Research Association, Chicago, IL.

Sundre, D. L. (1999). *Does examinee motivation moderate the relationship between test consequences and test performance?* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology, 29*(1), 6-26.

Taylor, C., Nahm, H., Quinn, M., Harms, M., Mulaik, J., & Mulaik, S. (1965). *Report of measurement and prediction of nursing performance, part I.* Salt Lake City, UT: University of Utah.

Tolman, E. C. (1925). Purpose and cognition: The determinants of animal learning. *Psychological Review, 32*, 285-297.

Vance, A., & Davidhizar, R. (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., Donald, E., & Widerquist, J. (1993). Predictors of success on the newest NCLEX-RN. *Western Journal of Nursing Research, 15*(5), 628 - 643.

Waterhouse, J., & Beeman, P. (2003). Predicting NCLEX-RN success: Can it be simplified? *Nursing Education Perspectives, 24*(1), 35 - 39.

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., Carrol, 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.

Weiner, B. (1986). *An attributional theory of motivation and emotion.* New York, NY: Springer-Verlag New York Inc.

Wendt, A. L. (2003). The NCLEX-RN examination: charting the course of nursing practice. *Nurse Educator, 28*(6), 276-280.

Wolf, L., & Smith, J. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*, 227-242.

Wolf, L., Smith, J., & Birnbaum, M. (1995). Consequence of performance, test motivation and mentally taxing items. *Applied Measurement in Education, 8*(4), 341 - 351.

Wyatt, A. (2002). *State suspends test-performance funding.* Retrieved May 19, 2005, from http://index.truman.edu/issues/20012002/story.asp?week=021402&SID=573

Yin, T., & Burger, C. (2003). Predictors of NCLEX-RN success of associate degree nursing graduates. *Nurse Educator, 28*(5), 232-236.

Yocum, C. (1998). *National Council of State Boards of Nursing: profiles of member boards*. New York: National Council of State Boards of Nursing.

APPENDIX A

Permission for Figure 1 An Attributional Theory of Motivation and Emotion

From: Essenpreis, Alice, Springer DE
To: tdrabbit56@earthlink.net
Cc: Rotter, Eugene, Springer, US
Sent: 8/8/2005 6:10:56 AM
Subject: AW: Lewis_Request for permission

Dear Ms. Lewis,

I am sorry for not getting earlier back to you.

We had some server problems, therefore, your e-mail might get lost, however, please find enclosed the permission grant:

With reference to your request (copy herewith) to re-use material
on which Springer controls the copyright, our permission is granted free of charge,
on the following conditions:
* it concerns original material which does not carry references to other sources,
* if material in question appears with credit to another source, authorization
from and reference to that source is required as well, and
permission is also obtained from the author (address is given on the imprint page or with the article);
* allows you non-exclusive reproduction rights throughout the world,
* permission includes use in an electronic form, on the condition that content is
- password protected,
- at Intranet or
- in CD-ROM/E-book;
* full credit (book/journal title, volume, year of publication, page,
chapter/article title, name(s) of author(s), figure number(s), original copyright notice)
is given to the publication in which the material was originally published by adding:
With kind permission of Springer Science and Business Media.
Permission free of charge does not prejudice any rights we might have
to charge for reproduction of our copyrighted material in the future.
With best regards,

–

Alice Essenpreis
Springer
Rights and Permissions

–
Tiergartenstrasse 17 | 69121 Heidelberg GERMANY
FAX: +49 6221 487 8223
Alice.Essenpreis@springer-sbm.com
www.springeronline.com/rights

APPENDIX B

Texas Woman's University Institutional Review Board Approval

TEXAS WOMAN'S UNIVERSITY
DENTON DALLAS HOUSTON
INSTITUTIONAL REVIEW BOARD
1130 John Freeman Blvd., Houston, Texas 77030 713/794-2074

MEMORANDUM

TO: Anne Young
Carolyn C. Lewis

FROM: IRB

DATE: August 8, 2005

SUBJECT: IRB Exempt Application

TITLE: Predictive accuracy of the HESI exit exam on NCLEX-RN pass rates and effect of progression policies on nursing student exam scores

APPENDIX C

Agency Permission

HESI

March 17, 2004

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 Carolyn Lewis

Dear Dr. Young:

This letter indicates my unconditional approval for Ms. Carolyn Lewis, 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. Lewis will be using test scores obtained from administrations of the HESI Exit Exam (E^2) during academic year 2001 - 2002 and responses from the HESI Annual Survey of nursing program administrators obtained in 2003 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. Lewis 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

Susan Morrison, PhD, RN
President