A

USING A HIERARCHICAL LOGISTIC REGRESSION MODEL TO ESTABLISH THE VALIDITY OF AN EXAMINATION WITH A DICHOTOMOUS CRITERION: POLICY IMPLICATIONS FOR NURSING EDUCATION

By

AMY ELIZABETH SCHMIDT

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy
The City University of New York

2000

UMI Number: 9959226

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

1/27/2000 David Rindskopf

Date Chair of Examining Committee

1/27/2000 Alan Gross

Date Executive Officer

David Rindskopf

Carol Kehr Tittle

Alan Gross

Supervisory Committee

The City University of New York

Abstract

USING A HIERARCHICAL LOGISTIC REGRESSION MODEL TO ESTABLISH THE VALIDITY OF AN EXAMINATION WITH A DICHOTOMOUS CRITERION: POLICY IMPLICATIONS FOR NURSING EDUCATION

by

Amy Elizabeth Schmidt

Advisor: Professor David Rindskopf

This paper presents a study that employed two-level hierarchical logistic regression models to establish the degree to which scores on the Diagnostic Readiness Test (DRT) predict success or failure on the National Council Licensure Examination for Registered Nurses (NCLEX-RN), and to determine how this relationship varies as a function of student-level and school-level variables. At the first level of these models, the predictors of NCLEX-RN performance included individual DRT and Nursing Pre-Admission Examination scores. At the second level, or school level, type of nursing program was used to explain how the relationship between DRT and NCLEX-RN varies from school to school.

For the final model, Empirical Bayes (EB) estimates of the parameters were obtained for each school, and were compared to the classical logistic regression coefficients. In addition, the overall predictive validity of the DRT was obtained, as well as the differential predictive validity for each type of nursing program. Results indicated that DRT scores were the only significant individual-level predictors of NCLEX-RN performance, and this relationship varied significantly by type of nursing program. In addition, the EB estimates were superior to the estimates obtained from the traditional logistic regression analysis, in that none of the EB slope estimates were negative.

Because this is the first systematic attempt to explain school-level variability in predictors of NCLEX-RN performance, policy implications for nursing education are also discussed and specific recommendations are presented. These recommendations address, for example, ways in which this type of hierarchical logistic regression analysis can benefit nursing education by identifying factors that influence NCLEX-RN performance for different types of schools. Implications for the further use of hierarchical analysis in psychometric applications are also discussed, as is the appropriateness of traditional validity models in evaluating the validity and utility of a diagnostic measure.

ACKNOWLEDGEMENTS

No work, however seemingly solitary a task, is ever produced in isolation. I'd like to thank the many people in my life that made it possible for me to pursue this degree and see it through (finally!) to completion. First of all, profound thanks are due to my dissertation supervisor, David Rindskopf, who was always there when I needed him to provide guidance, encouragement, support, and best of all, many a good laugh. I'd also like to thank the other members of my committee, Alan Gross and Carol Kehr Tittle, whom I've admired as professors, scholars, and mentors. All three of these people also need to be acknowledged for the fine job they did in ushering me from "student-hood" to professionalism.

I owe a huge debt of gratitude to my family – to my parents, Helen and Werner; my sisters, Caroline and Jennifer; and my brother-in-law, Robert; for providing me with the foundation that I needed to succeed and the encouragement to persevere through the difficulties that life presented along the way. To the next generation – to my nephews Brian, Brendan, Rob and Conor, and my godchildren, Stephen, Kiersten and Kirsten - I wish you the same unconditional love and support that I received as you journey through life (and knowing your parents, you'll get it!). To my aunts, Claire and Betty, and my cousins, all of whom suffered great loss within the last couple of years, I want to say that I admire your strength and thank you for your support, even as you have gone through difficult times. I'd also like to acknowledge those who have gone on ahead: my grandparents Lisa, Alfred and Clara; my uncles Russell and Bill; my aunt Honey; and my cousin Alan - I wish you were all still here, but I know that you're celebrating along with the rest of the family!

Many thanks are due to my friends, particularly those who have traveled a similar path. To Kathy Van Aken, Stewart Schuele, Eric Rodriguez, James Pratt, Dorinda Welle, Tracy Busch, Daphne Greenberg, Perry Halkitis, Debbie Hecht, Julia Leahy, Anne Harvey, Carolyn Crone, Cathy Wendler, Ida Lawrence, Susan Phillips, Anita O'Brien, Jeff Burkeen, Tricia Callender, Bob Borden, Jill Thistlethwaite, Catherine Masterson, and the St. Peter's choir – thanks for all the sympathy, encouragement and laughs.

Last, but certainly not least, I'd like to acknowledge the support of my employer, the College Board, for providing the most valuable resource of all – time! I'd also like to thank my College Board R&D colleagues, Wayne Camara, Thanos Patelis, Marlene Dunham, Howard Everson, Robert Majoros, Jeri Ann Cook, Gem Lucas, Avis Hudson, Julie Nathan, Tammy Austin, Paula Kuebler and Josh Philips for putting up with my craziness as I finished up, with special thanks to Avis for all her administrative help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Establishing the predictive validity of an instrument is at issue when the purpose of the instrument is to estimate performance on a criterion measure (Anastasi, 1988). The method most frequently used in establishing evidence of such validity is to correlate scores on the predictor test with scores on the criterion variable. In cases where the criterion variable is dichotomously scored, an appropriate model for predicting this criterion from the continuously scored predictor would be the logistic regression model (Hosmer & Lemeshow, 1989). However, this model may not be adequate if the data involved have a nested, or multilevel structure. When subjects are nested within a second level of analysis, and second-level variables are known to affect performance on the criterion measure, it is not possible, using a traditional logistic regression model, to adequately separate individual variation from second-level variation (Bryk & Raudenbush, 1992; Kreft & de Leeuw, 1998). Multilevel models, or hierarchical models (Lindley & Smith, 1972) allow this separation.

Hierarchical models are a generalization of traditional regression models applied to nested data. A common example of this type of nesting in educational settings is that of students nested within schools, with student-level data forming the first level, and school-level data forming the second level. The model can be extended further to a third level, for example, schools within districts or states.

Lindley and Smith (1972) introduced the term "hierarchical linear model" in a paper describing a reanalysis of linear models using Bayesian methods. For the first time, a model was available which allowed for an analytical representation of the hierarchical character of much of the data that is collected in the social sciences.

In addition to the nesting problem in hierarchical data structures, it is very often the case that the number of observations at the first level per second-level unit (e.g., the number of students per school) is too small to permit accurate estimation of the within-second level (e.g., within-school) parameters. The procedure commonly used in hierarchical models to estimate model parameters is empirical Bayes (EB), which allows for the pooling of data across the second level units (Braun, 1989). These estimates are particularly useful in psychometric contexts, such as establishing the validity of a test. Empirical Bayes approaches to establishing test validity have been developed and applied successfully in a number of studies in which test scores are used to predict a continuous criterion (Braun, Jones, Rubin, & Thayer, 1983; Rubin, 1980; Zwick, 1993). In the context of linear models, EB estimates have generally been shown to be superior to ordinary least squares estimates. In addition, EB models can incorporate second-level variables so that they may be treated as covariates, providing estimates of within-school parameters that reflect the effects of between-school variation.

While a body of research exists that examines prediction of a continuous criterion using EB techniques, the literature is lacking in studies that explore prediction of a dichotomous criterion. In psychometric contexts, it is quite common to report "Pass/Fail" rather than a continuous score in licensure and certification examinations. Very often, these data take on a nested structure: generally, a pool of candidates for licensure in some profession have completed their course work at different schools, and some schools may have graduated only a small number of students in a given time period. Using an EB approach with logistic regression models should result in more accurate coefficient estimates for each school; estimates that take into account variation due to school-level predictors as well as variation due to individual-level predictors. It would seem then to be of value to use hierarchical logistic regression models to evaluate the predictive validity of a test that purports to predict performance on this type of licensure or certification examination.

This dissertation presents the results of a validity study that examines the degree to which scores on the Diagnostic Readiness Test predict success or failure on the National Council Licensure Examination for Registered Nurses. A preliminary study that utilized an approximation to a hierarchical logistic regression analysis concluded that there was significant school-level variation in the distribution of slopes and intercepts obtained from predicting success/failure on the NCLEX-RN from DRT scores (Schmidt, 1997). This follow-up study is being conducted to explore further this relationship by employing a hierarchical logistic regression model, which may well result in more accurate coefficient estimates for each school, as well as allow for a separation of error variance from true variance at the school level, which was not possible with the approximation method (Schmidt, 1997).

Researchers and theorists in test development have been challenging the notion of treating validity as though there were three distinct types - construct, content, and criterion – for the last several decades. These investigators have argued for a unified concept of validity in which the concept is expanded to a more comprehensive theory of construct validity, with particular emphasis on the consequential aspects of score interpretation (Cronbach, 1988; Messick, 1989; Messick, 1995). However, a predictive validity study does have a place in this widely accepted reconceptualization. Indeed, if we take Messick's definition of validity as "... nothing less than an evaluative summary of both the evidence for and the actual - as well as potential - consequences of score interpretation and use" (Messick, 1995, page 742), the results of a predictive validity study can certainly add to this evidence as long as it is accompanied by a critical discussion of how these results are interpreted and used by the communities that are served by the results. Therefore, because this is the first systematic attempt to model the school-level variability in predictors of NCLEX-RN performance, policy implications for nursing education are also discussed and specific recommendations addressing ways in which this type of analysis can benefit nursing education are also presented.

CHAPTER II: REVIEW OF THE LITERATURE

A. Predictive Validity

Criterion-related validation procedures have been traditionally defined as "procedures (that) indicate the effectiveness of a test in predicting an individual's performance in specified activities" (Anastasi, 1988, page 145). Therefore, performance on a test is examined with particular emphasis on its relationship to an external criterion measure, which should be an independent measure of the construct that the test is designed to predict. In establishing the criterion-related validity of a test, the test may be administered at the same time that the criterion measure is assessed, or at the time that information on the criterion measure is available. Alternatively, the test may be administered first, and the criterion assessed at a later time. The first type of criterion-related validity is referred to as concurrent validity, while the latter procedure is referred to predictive validity (Anastasi, 1988).

A common method used to establish the predictive validity of a test is to correlate the scores on the test with scores on the criterion measure. In the specialized case of a dichotomously scored criterion measure, an appropriate model for predicting this criterion from the continuously scored predictor would be the logistic regression model (Hosmer & Lemeshow, 1989). According to the *Standards for educational and psychological testing* (APA, AERA, NCME, 1999), predictive validity studies are considered critical in the development of assessment instruments whose results are intended to inform decision-making processes, such as employment decisions, clinical diagnoses, and educational assessment. These types of decision-making processes are often modeled with a dichotomous dependent variable; for example, the decision to hire or not hire a candidate (Raju, Steinhaus, Edwards, & DeLessio, 1993), the assessment of whether a client should be diagnosed with a particular disorder or not (Records & Tomblin, 1994), or the determination of what factors predict whether a student will successfully complete an educational program (House & Johnson, 1993; Spahr, 1995). Because the present study is set in the educational arena, the remainder of this section of the literature review will be devoted to discussion of those studies that established predictive validity for educational purposes.

Predictive Validity Studies Conducted at One School

Spahr (1995) examined the relationship of entering grade point average (GPA) and grades in prerequisite support courses in algebra, biology, and chemistry to graduation in the college's nursing program at Morton College in Illinois. Spahr obtained a sample consisting of 255 students admitted to the nursing program in 1990, 1991, and 1992. A mixture of analytic techniques was employed: multiple regression was used to predict graduation status from all predictors, while logistic regression was performed to estimate a student's odds of graduating from the nursing program based solely on GPA. The results of the multiple regression indicated that the proportion of variance in graduation status that was explained by the significant predictors (GPA, grades in biology, grades in chemistry) was very small, and the results of the logistic regression procedure indicated only a modest increase in the odds of graduating as GPA increased. Spahr interpreted these results as indicating that the effects of the predictors were inconsequential in predicting the odds of graduation. However, the results Spahr obtained are hardly surprising considering the incorrect use of multiple regression with a dichotomous criterion, graduation status. The prediction of a dichotomous criterion becomes more difficult than the case wherein the criterion is continuous because of the loss of information on the criterion measure, so the small proportion of variance in graduation status that was accounted for by the analysis should have been anticipated with the use of multiple regression. Even though the use of multiple regression techniques may produce reasonable estimates when the dichotomous outcome probability is between .2 and .8, when the probabilities are at the extreme, the R^2 that results from a multiple regression analysis provides a low estimate of the true relationship between the predictors and the criterion.

House and Johnson (1993) investigated the predictive validity of Graduate Record Examination (GRE) scores and academic background variables for graduate degree completion in psychology among 250 graduate students at a large Midwestern university. The dichotomous criterion measure was whether or not these students obtained the Ph.D. degree. Stepwise logistic regression procedures were used to analyze the relative contribution of each GRE score and academic background variable toward explaining subsequent degree completion. Interestingly, results indicated that the predictor variables did not predict similarly for all students - GRE verbal scores were the strongest significant predictor of degree completion for students in professional psychology areas, but the weakest significant predictor in general/experimental psychology programs. In this study, logistic regression was properly employed to examine the relationship between the predictors and graduation status, but the results may not be generalized to other populations because the investigators only used students from one university, as did Spahr (1995). Even if these investigators had expanded their studies to other university populations, the use of the logistic regression procedure alone could not have properly modeled the effects of school-level characteristics upon the relationship between the predictors and the criterion.

Predictive Validity Studies Conducted at Multiple Schools

In recent years, some researchers have been exploring alternate methods of analysis to account for second-level variability when presented with the problem of establishing the predictive validity of an examination when the data are obtained from individuals across many schools. Although none of the studies that were conducted to specifically address this problem used a dichotomously scored criterion, a review of these studies will help to shed light of the current 'state of the art' in predictive validity.

A common problem that is often encountered when validity studies are conducted with data from individuals across schools occurs when individual schools require their own validity coefficient. In traditional approaches to analyzing this kind of data, least-squares estimates of the coefficients are obtained, but these coefficients can vary widely between schools and from year to year. One of the first researchers to develop an alternative method of estimating these within-school validity estimates was Rubin (1980), who examined the ability of undergraduate grade point average and scores on the Law School Aptitude Test (LSAT) to predict first-year grade average in law school across 82 law schools. Rubin found that EB estimates that used information from all the law schools that participated in the study were more stable over time than the least-squares estimates, and predicted a student's performance more accurately.

Another problem that arises in multi-school studies is when the number of students in each school varies widely, and some school have too few students to develop stable within-school estimates of the validity coefficient. Braun, Jones, Rubin, and Thayer (1993) addressed this problem in a study that examined the ability of scores on the verbal and quantitative portion of the Graduate Management Admissions Test (GMAT) and undergraduate grade point average to predict first year grade average in business school. The problem of interest in this study concerned establishing separate regressions for Black students in each school. Unfortunately, of the 59 business schools available for analyses, only 24 had at least four Black students, and Black students represented only 4% of the total sample. This paucity of minority data made it impossible to estimate reliable least squares regression coefficients for each school. The investigators were able to fit the data to an EB model and found that the EB estimators were more predictive of first year grade average than the least squares estimates.

Using the techniques developed by Braun, et al. (1983), Zwick (1993) examined the validity of the Graduate Management Admission Test (GMAT) as a predictor of grades in graduate business programs leading to a doctorate. Again, because of the small sample sizes within certain schools, Zwick used an EB approach to estimating the parameters of four different models. In an extension of Rubin (1980) and Braun, et al.'s (1983) work, Zwick included a school-level covariate, school mean on GMAT total score as a measure of school selectivity, in three of the four models. Zwick found, however, that the simpler model that did not include the covariate was preferable. She hypothesized that this was due to the lack of variability in the school selectivity measure, which reflected the higher selectivity of doctoral programs as opposed to the MBA programs, used in the Braun, et al. (1983) study.

As beneficial as EB methods have been to the study of predictive validity, there has been no similar application of these methods to the case in which the criterion is dichotomously scored. However, EB estimators have been used in hierarchical logistic regression models in survey research. An example is a study of contraceptive use in developing nations (Wong & Mason, 1985). In this study, both classical and Bayesian logistic regression coefficients were developed in predicting whether women ever used contraception, based on a series of individual-level predictors, such as the woman's education, and country-level predictors, such as gross national product. While the classical and Bayesian logistic regression coefficients were not compared in terms of their stability or accuracy, the study's results suggested that the Bayes estimates were more interpretable than the classical estimates.

There still remains a void in the literature in terms of studies that use EB estimates in hierarchical logistic regression models to establish the predictive validity of an examination. This study presents the results of a validity study that examines the degree to which scores on the Diagnostic Readiness Test (DRT) predict success or failure on the National Council Licensure Examination for Registered Nurses (NCLEX-RN). As the DRT was developed for the purpose of predicting performance on the NCLEX-RN, as well as to provide feedback to the student before the licensure examination is administered, it is critical that the predictive validity of this examination be established with a large, representative sample, in keeping with the *Standards for educational and psychological testing* (APA, AERA, NCME, 1999). This dissertation addresses this need by examining the predictive validity of the continuously scored DRT in predicting success or failure on the NCLEX-RN using a national sample. In addition, as mentioned previously, in keeping with the unified concept of validity set out by Cronbach (1988) and Messick (1989; 1995), the results of this study will contribute to the body of evidence for the construct validity of the DRT as a predictor of NCLEX success by setting the results obtained in this study in the educational and political context in which the users of this examination function. Only then can reasonable and thoughtful policy recommendations be developed for nursing education. We now examine the history of previous attempts to predict performance on the NCLEX-RN in order to understand the current state of affairs in nursing education as it pertains to preparing students for licensure.

B. The NCLEX-RN

The National Council Licensure Examination for Registered Nurses is an examination required for all nursing school graduates in order for the nurse to become licensed to practice. Students of nursing may sit for the licensure examination after completion of course work in one of three different types of nursing programs: diploma programs, associate degree programs, or baccalaureate degree programs. These three programs are different in philosophy and course work covered, yet each program graduates students who are licensed to practice under the title of registered nurse. The diploma program, the oldest type of nursing program, is based on an apprenticeship model and is usually hospital-based. It has been argued that the diploma program provides early and consistent clinical exposure, and earlier socialization into the role of nurse, therefore providing the nurse with a more realistic comprehension of the role (Riffle, Lamberth, Moine, & Fielding, 1985).

In 1965, the American Nursing Association (ANA) proposed the baccalaureate degree as the professional degree for entry into practice, and there has been intense debate on this topic in the nursing community ever since. More recently, in the wake of the nursing shortage of the mid- to late-1980's, many critics have come down firmly on the side of baccalaureate education, arguing that it provides nurses with a greater theoretical basis for practice that enables the nurse to deal effectively with the rapidly changing and complex modern health care system (Duffey, 1990; Fitzpatrick & Modly, 1990; Oermann, 1991; Sakalys & Watson, 1986; Williamson, 1983). These authors have also argued for more clearly differentiated levels of nursing practice and related health care roles (i.e., practical nurse, physician's assistant, etc.) in order to enhance the status of the registered nurse as a professional, and so argue for the higher degree as the minimum for entry into professional practice.

In view of this debate, it is of interest to examine how well students from each of these three types of programs perform on the NCLEX-RN. Typically, studies that have examined differences in pass rates between the three groups have found that baccalaureate graduates have the lowest pass rates, while associate degree graduates have the highest (Waterhouse, Carroll, & Beekman, 1993). While this may seem surprising, a review of the NCLEX-RN test content helps to provide a reasonable hypothesis for this difference. The NCLEX-RN test plan, developed by the National Council of State Boards of Nursing, is based on a job analysis that is conducted every three years. This job analysis surveys newly-licensed nurses (i.e., nurses who have had their license for approximately six months) regarding the types of activities they perform on a regular basis and the criticality of these activities, as well as the knowledge they need to perform these activities effectively (Chornick & Yocom, 1995). As might be expected, newly-licensed nurses perform lower-level duties than do experienced nurses and are not required to call upon higher-order knowledge and thinking skills until they have more clinical experience, and the content of the NCLEX-RN represents this entry-level knowledge and skills. Moreover, the pass rate of the NCLEX-RN is set to ensure minimal competency for safe practice (Chornick & Yocom, 1995). Therefore, the knowledge and skills that baccalaureate graduates pride themselves on, while they may certainly be useful later on in their nursing careers as they assume leadership positions, may hinder rather than help baccalaureate graduates enter nursing practice.

C. Predicting Performance on the NCLEX-RN

Interest in predicting performance on the NCLEX-RN has waxed and waned over the years. During the recent shortage of qualified nurses that reached its peak in the 1980's, schools of nursing became particularly concerned about their graduates' NCLEX-RN performance, and conducted studies to explore significant predictors in order to re-examine their educational processes and curriculum, particularly in terms of preparing candidates for the licensure examination (Waterhouse, et al., 1993). Clearly, it is in a school's best interest to produce graduates who can become licensed as soon as possible: this enhances a school's reputation and may eventually affect its enrollment. Additionally, schools that are undergoing accreditation review are particularly concerned with assessing cumulative student achievement prior to graduation. These schools seek out valid and reliable measures of achievement that provide an indication of a student's chances of passing the NCLEX-RN in order to document their effectiveness (Beitz, 1994). Therefore, providing schools with a properly validated examination that can accurately predict NCLEX-RN performance will enable schools to be positioned better to identify students who require remedial instruction prior to the licensure examination. These schools will also be better prepared to demonstrate to accreditation review boards their effectiveness in covering the curriculum required for successful performance on the licensure examination.

There are many studies that examine possible predictors using data from NCLEX-RN administrations prior to 1988, some of which use only individual-level predictors, such as grade point average, and some of which use a combination of individual- and school-level predictors. Before 1988, the NCLEX-RN scores were reported as a continuous measure, with 1600 set as the pass score. Since then, with the advent of a new test plan based on the 1985 job analysis, NCLEX-RN scores are reported as "pass/fail" only (Chornick & Yocom, 1995; Matassarin-Jacobs, 1989), making predictive analyses technically more difficult because of the attenuation of the validity coefficient due to the loss of information on the criterion measure, NCLEX-RN scores (Drake & Michael, 1995). Unfortunately, none of the post-1988 studies include school-level predictors, perhaps because of the generally accepted notion that each nursing program has its own optimal combination of unique predictors (Foti & DeYong, 1991).

Studies with Individual-Level Predictors

A recent review of studies that examined various predictors of pre-1988 NCLEX-RN performance conducted by Campbell and Dickson (1996) concentrated solely on studies that predicted the performance of baccalaureate graduates. Campbell and Dickson reviewed 47 studies and found that scores on pre-admissions tests, such as the SAT or ACT, overall grade point average and scores on the National League for Nursing's Baccalaureate Comprehensive Examination were generally significant predictors of NCLEX-RN success.

However, the authors point out that all of the studies used relatively small samples of convenience and emphasized the need for collaborative efforts across schools. Other studies examined only associate degree graduates (Perez, 1977; Reed & Feldhusen, 1972) and found strong relationships between pre-admissions tests such as the SAT and ACT and NCLEX-RN performance.

One study examined the relationship between a set of predictors and pre-1988 NCLEX-RN scores across two types of programs within the same nursing school. However, the primary purpose of Gross, Takazawa, and Rose's (1987) research was to evaluate the impact of nursing education on student's critical thinking skills. As a supplementary analysis, scores on the Watson-Glaser Critical Thinking Appraisal (WG) and the NLN's Pre-Admission Test were related to grade point average and NCLEX-RN scores. The sample consisted of 60 associate degree students and 60 baccalaureate degree students at the University of Hawaii School of Nursing. Results indicated that the WG scores were predictive of both grade point average and NCLEX-RN performance for baccalaureate students, but not for associate degree students. Gross, et al. (1987) did not interpret this finding, however, because differential prediction of NCLEX-RN performance by program type was not the main thrust of their analyses. This research, which is one of the few to examine prediction across program type, was also limited by the small, localized sample.

Only two published studies and one unpublished doctoral dissertation exist that examine predictors of post-1988 NCLEX-RN performance. Waterhouse, et al. (1993) used a sample of available baccalaureate graduates at the University of Delaware. Using a discriminant analysis these authors found that SAT verbal scores and end-of-second-year grade point average were significant predictors of success or failure on the NCLEX-RN. The discriminant function also successfully predicted 91% of the passing graduates; however, this information must be interpreted in light of the fact that the sample had a pass rate of 83%.

In another study utilizing discriminant analysis, Akers (1992) found that there were a different set of predictors of NCLEX-RN performance for the associate degree students than there were for the baccalaureate students. Using a sample of 1,894 students enrolled in either a baccalaureate program or an associate degree program in the state of Mississippi, Akers conducted two separate series of stepwise discriminant analyses: one for each of the program types. Akers found that for the baccalaureate students, ethnic origin, grade point average, failing a nursing course one or more times, and scores on the ACT, Diagnostic Readiness Test and the NLN's Pre-Admission Test were predictive of NCLEX-RN performance, while predictors for associate degree students included age, ethnicity, grade point average, failing a nursing course one or more times, scores on the Diagnostic Readiness Test, and scores on a variety of

NLN achievement tests. However, Akers did not attempt to compare the relative predictive power of the common significant predictors, such as grade point average and Diagnostic Readiness Test scores, across the program types. Even if she had, it would be difficult to interpret those differences without a sense of how much of the between-school variability in NCLEX-RN performance is due to a school-level variable like program type.

Drake and Michael (1995) took a somewhat different approach with a sample of associate degree graduates at Fresno City College. They examined the predictive power of grades in specific nursing and science courses, as well as various grade point averages, in predicting success or failure on the NCLEX-RN. Although these authors acknowledged the problems in attenuation of the validity coefficient due to the dichotomous criterion measure, they used a series of Pearson correlation coefficients to determine the relationship between each of the predictors (24 in all) and NCLEX-RN performance. Drake and Michael found that the grade point average for eight nursing theory courses was the strongest predictor of NCLEX-RN performance.

Studies with School-level Predictors

Very little work exists wherein the primary purpose of the research is to examines predictors of NCLEX-RN performance across types of nursing programs, and the work that does exist consists primarily of unpublished doctoral dissertations, examining different types of programs within the same state. Two of these unpublished studies will be presented here in order to illustrate the kinds of analyses performed across nursing programs utilizing school-level predictors.

Campbell (1988) surveyed 342 faculty members from 60 schools of nursing located in Ohio, representing all three levels of nursing preparation. Campbell obtained information on such school-level predictors as student-faculty ratio, instructional activities, curriculum content, and educational environment of the nursing programs from the faculty members and from the Ohio State Board of Nursing Education and Nurse Registration, and used these to predict the individual-level scores on the continuously scored 1987 NCLEX-RN. Analyses were conducted separately for each type of program, and Campbell found that different sets of independent variables were significant predictors of NCLEX-RN performance for each of the three types of programs. There are obvious methodological flaws with this study. In order to obtain school-level measures on each of the variables of interest, such as educational environment, Campbell averaged the responses of the faculty associated with each school, despite the fact that the survey return rate and faculty sample size varied by school. She then used the school as the unit of analysis, with proportion of students passing the NCLEX-RN as the dependent variable. If Campbell had structured the data set in a hierarchical data fashion, employing faculty as the first level unit of analysis nested within nursing schools, she could have used the entire data set in one analysis. In this manner, type of nursing program could have been used as a second-level predictor, and she would not have had to conduct three separate sets of analyses.

Using a different approach, Karns (1988) examined the records of the Pennsylvania State Board of Nursing to investigate possible trends in NCLEX-RN performance as a function of program type from 1969 to 1985. Using Chi-square analyses, Karns found that there were significant patterns and trends in the pass rates of students from different types of programs. As the number of diploma programs in Pennsylvania decreased during this entire time period, the NCLEX-RN pass rate decreased until 1980, and then rose. In contrast, as the number of associate degree programs increased, the NCLEX-RN pass rate increased. The baccalaureate programs had the highest success rates during the first four years while the number of baccalaureate programs remained stable, but then, as the number of baccalaureate programs increased, the NCLEX-RN pass rate decreased. Karns interpreted these trends in terms of the costs and benefits associated with either decreasing or increasing enrollment, but did not address the ways in which different curriculums evolved during this period as the three types of programs became conceptually distinct (Riffle, et al., 1985).

It is clear from this review that, although efforts have been made to identify predictors of NCLEX-RN performance, these efforts have been somewhat uneven, particularly in the post-1988 data, both conceptually and methodologically. Small, localized samples have been used with graduates from one type of program at a time. Inappropriate statistical techniques have been employed, such as Drake and Michael's use of Pearson correlation coefficients (1995). No systematic effort has ever been made to take into account school-level variables that may affect success rates on the NCLEX-RN even though some authors acknowledge that the strength of various predictors may vary by school (Foti & DeYong, 1991). Certainly, nothing as sophisticated as the use of hierarchical models has been attempted to model the relationship between possible predictors of NCLEX-RN performance, both at the individual and school level, and NCLEX-RN scores. The results of this dissertation address these concerns by modeling the hierarchical nature of the data in a logistic regression framework.

D. Hierarchical Linear Models

The use of hierarchical linear models has been increasing in recent years, and these models have been applied to a wide variety of problems (Draper, 1995; Goldstein & McDonald, 1988; Rubin, 1989), including analysis of change in repeated measures designs (Bryk & Raudenbush, 1987; Rogosa & Saner, 1995), analysis of nested designs (Lee & Bryk, 1989; Mason, Wong, & Entwistle, 1983; Vancouver, Millsap, & Peters, 1994), meta-analysis (Draper, Gaver, Goel, Greenhouse, Hedges, Morris, Tucker, & Waterneaux, 1993; Raudenbush, 1984) and, as previously mentioned, psychometric analyses (Bergstrom, Gershon, &

Lunz, 1994; Braun, Jones, Rubin, & Thayer, 1983; Rubin, 1980; Rubin, 1989; Zwick, 1993;).

Hierarchical models are a generalization of traditional general linear models applied to nested data when the nesting occurs within random factors. In experimental designs, where the subjects are nested within fixed factors, the analysis of variance procedure (ANOVA) is an appropriate analytic technique, and no special modeling of the hierarchical structure of the data is necessary. An example of this might be a design where subjects are randomly assigned to four different clinical treatment modalities, and these specific treatments are to be compared for their efficacy. However, consider the same example with a twist: twenty-five therapists are administering the treatments, and subjects are nested within therapist. In this case, therapist is a random factor. When this kind of nesting occurs, hierarchical modeling is necessary to appropriately model the second-level random variation.

A common example of this type of nesting in educational settings is that of students nested within schools, with student-level data forming the first level, and school-level data forming the second level. The model can be extended further to a third level, for example, schools within districts or states. Lindley and Smith (1972) introduced the term hierarchical linear model in a paper describing a reanalysis of linear models using Bayesian methods. For the first time, a model was available that allowed for an analytical representation of the hierarchical character of much of the data that is collected in the social sciences.

It has been argued that using hierarchical models rather than traditional models for the analysis of nested data has distinct advantages. Investigators have maintained that hierarchical models provide a better model of nested data and therefore allow the formulation and testing of hypotheses that could never be properly tested with traditional models (Bryk & Raudenbush, 1992; Draper, 1995; Goldstein, 1987). As was explored in the section on predictive validity, hierarchical models can produce improved estimation of individual effects (e.g., Braun, Jones, Rubin, & Thayer, 1983; Rubin, 1980). Hierarchical models can also be used to model cross-level effects (e.g., Bryk & Raudenbush, 1987; Mason, et al., 1983; Raudenbush, 1993). For example, in the Mason, et al. (1983) study, the investigators were interested in exploring the relationship between maternal education level, urban versus rural residence, and fertility rates across fifteen countries. They found that while higher maternal education was associated with lower fertility in each of the fifteen countries, the results for urban versus rural residence was not as clear-cut. When country-level variables were taken into account, however, it was seen that gross national product and availability of family planning services could explain the difference in the urban and rural fertility rates, which varied from country to country.

Another problem with nested data that was unable to be effectively addressed until the advent of hierarchical models is the problem of separating variance into between-second level and within-second level components. For example, in a 1994 study, Bergstrom, Gershon and Lunz examined examinee response times to items from a computerized adaptive test. The investigators hypothesized that there would be greater variation in response time across candidates than within candidates, and that demographic characteristics such as age, gender, ethnicity, and personality variables such as level of test anxiety could explain a significant proportion of this variation. They found, however, that variance within persons was eight times greater than variance between persons, and that this variance could be explained by item characteristics, such as length, position of the correct answer, and the use of figures in the item. The hierarchical linear model that was employed allowed the partitioning of the variance so that this relationship could be revealed.

Although the use of hierarchical models has been extended to analyses of variables with nonnormal distributions, as will be presented in this dissertation, the bulk of the research conducted with these models has been in the realm of predicting a continuous dependent measure. A typical hierarchical linear model, or random coefficients model, is presented below.

Consider the typical linear regression equation that models the relationship between an individual-level predictor and an outcome variable for students who are nested within schools. Within each school, the model can be represented:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$$

Where:

Y_{ij} = The value on an outcome variable for student i in school j

β_{0j} = The intercept for school j, or the expected value of the outcome variable for an individual for whom the value of the predictor variable is zero

β_{1j} = The slope, or the expected change in the outcome variable for every unit change in the predictor variable

X_{ij} = The value on a predictor variable for student i in school j

r_{ij} = The error, or residual, term, representing a unique effect associated with student i in school j

In these types of models, it is usually assumed that r_{ij} is normally distributed with homogeneous variance σ^2 across schools (Bryk & Raudenbush, 1992). Note that if this equation were developed separately for each school, each school would be permitted to have its own intercept and slope, β_{0j} and β_{1j}. It is usually assumed that the intercept and slope have a bivariate normal distribution across the population of schools (Bryk & Raudenbush, 1992). Let:

$E(\beta_{0j}) = \gamma_{00}$ $\quad$ $Var(\beta_{0j}) = \tau_{00}$

$E(\beta_{1j}) = \gamma_{10}$ $\quad$ $Var(\beta_{1j}) = \tau_{11}$

$Cov(\beta_{0j}, \beta_{1j}) = \tau_{01}$

Where:

γ_{00} = the average conditional school mean for the population of schools for $X_{ij} = 0$

τ_{00} = the population variance among the conditional school means

γ_{10} = the average slope for the population

τ_{11} = the population variance among the slopes

τ_{01} = the population covariance between the slopes and intercepts

Second-level predictors may account for the variability in the slopes and intercepts. This can be represented by the following example using one level-2 predictor:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

and

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$$

Where:

γ_{00} = the mean of the outcome variable for a school for which the value of the second-level predictor variable is zero

γ_{01} = the mean expected change in the outcome variable for every unit change in the second-level predictor variable

γ_{10} = the mean slope for a school for which the value of the second-level predictor is zero

γ_{11} = the mean expected change in the slope for every unit change in the second-level predictor variable

u_{0j} = the unique effect of school j on mean outcome variable, conditioning on W_j

u_{1j} = the unique effect of school j on the slope, conditioning on W_j

W_j = the value of the second-level predictor for school j

Substituting the second-level equations into the first-level equations, the combined model becomes:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}X_{ij} + \gamma_{11}W_j X_{ij} + u_{1j}X_{ij} + u_{0j} + r_{ij}$$

This combined model presents some difficulties in terms of parameter estimation. In ordinary least squares estimation, it is assumed that the errors are independent, randomly distributed, and have constant variance. Compare these assumptions with the error term in the random coefficients model given above. As can be seen, the error term, $u_{1j}X_{ij} + u_{0j} + r_{ij}$, is dependent within each school, since the u_{0j} and u_{1j} are common to every student within a particular school. In addition, the errors have unequal variances, because $u_{0j} + u_{1j}X_{ij}$ varies across schools, and X_{ij} varies across students. Therefore, standard estimation procedures cannot be used with these models. We now examine how EB estimates are used to obtain accurate parameter estimates.

E. Empirical Bayes (EB) Estimates

The logic behind using EB estimates can seem counter-intuitive, as illustrated in Efron & Morris' (1975) article on what they call 'Stein's paradox' - the finding that an estimator other than the mean is superior to the mean when many means are being estimated. This estimator, referred to as the James-Stein (or EB) estimator, 'shrinks' each mean toward the grand mean. The amount of shrinkage depends on the standard error of the quantity estimated, which in turn depends on the sample size. Therefore, if a particular mean is based on only a few observations, its standard deviation will be large, and the amount of shrinkage toward the grand mean will also be large. Conversely, when a mean is calculated from many observations, it may be regarded as a more stable estimate. Its standard error will be smaller, and the amount of shrinkage towards the grand mean will be smaller. This approach can be used as an alternative to least squares estimation in estimating regression coefficients as well, and indeed, EB estimates have been shown to be superior in terms of their stability across time, as well as in their practical uses with unbalanced, or nested data (Braun, 1989). In hierarchical linear models, the goal is to estimate many β's across the second level units, using information obtained not only at the individual level, but information from the second level as well. As an example, consider the simple case with no level-one or level-two predictors. Using Bryk and Raudenbush's (1992) notation, the level-one equation unit can be written:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

The parameter β_{0j} can be expressed as:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

The EB estimator is then a weighted estimate of the sample mean, $Y_{\cdot j}$, and the grand mean, γ_{00}, and is given by:

$$\beta_{0j}{}^{*} = \lambda_j Y_{\cdot j} + (1 - \lambda_j)\, \gamma_{00}$$

where:

$$\lambda_j = \mathrm{Var}(\beta_{0j}) \,/\, \mathrm{Var}\,(Y_{\cdot j})$$

It can be seen from this equation that the sample mean is "shrunken" toward the conditional mean. (It should be noted that the conditional mean is near the grand mean when the level-two predictor does not explain a substantial proportion of the variance in β_{0j}.) When λ_j is large, as happens in the case when the sample mean is a highly reliable estimate of β_{0j}, there is less shrinkage of the sample mean towards the conditional grand mean. When λ_j is small, as happens in the case when the sample mean is not a reliable estimate of β_{0j} (e.g., when the sample size at a particular second-level unit is small), there is a greater amount of shrinkage of the sample mean towards the conditional grand mean. These weighted averages are generally considered superior to ordinary least square estimates in that no other estimate of β_{0j} has a smaller expected mean square error (Lindley & Smith, 1972).

F. The Hierarchical Logistic Regression Model

A hierarchical logistic regression model using an EB estimation procedure was first proposed in 1985 by Wong and Mason in their continuing analysis of the World Fertility Data. In this model, Wong and Mason attempted to predict whether or not a woman ever used modern contraception methods (EVER) from both individual level variables, such as number of years of formal schooling for the wife (WED) and her type of residence during childhood (URBC; coded as 1 for urban and 0 for rural), and country level variables, such as gross national product (GNP) and effectiveness of family planning programming (FPE). Wong and Mason posited the following two-stage hierarchical logistic regression model. At the first level, the usual logistic regression model is applied to first level observations within each of j contexts:

$$\text{Logit}(\text{EVER}_{ij}) = \beta_{0j} + \beta_{1j}\text{WED}_{ij} + \beta_{2j}\text{URBC}_{ij}$$

Where:

$\text{Logit}(\text{EVER}_{ij})$ = log-odds of using contraception for woman i in country j [the log-odds is defined as the natural logarithm of the probability of ever using contraception divided by 1-(the probability of ever using contraception); $\text{Logit}(\text{Ever}_{ij}) = \ln\{p(\text{Ever}_{ij})/1-p(\text{Ever}_{ij})\}$]

WED_{ij} = Wife's education, in number of years of formal schooling, for woman i in country j

$URBC_{ij}$ = an indicator variable to code place of residence; 1 if urban, zero if rural, for woman i in country j

β_{0j} = intercept term for country j, or the expected value of the log-odds of using contraception for a rural individuals (URBC = 0) for whom the value of WED (wife's education) is zero

β_{1j} = slope for WED value in country j, or the expected change in the log-odds of using contraception for every unit change in WED (wife's education)

β_{2j} = slope for URBC in country j, or the expected change in the log-odds of using contraception between urban and rural dwellers

Again, as in the hierarchical linear model, it was assumed that the intercept and slope have a bivariate normal distribution and were free to vary across the population of countries. A more general model that incorporates country-level variables, such as GNP (gross national product) and FPE (family planning program), and the interaction between them, was developed:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}GNP_j + \gamma_{02}FPE_j + \gamma_{03}GNP_j{*}FPE_j + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}GNP_j + \gamma_{12}FPE_j + \gamma_{13}GNP_j{*}FPE_j + \mu_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}GNP_j + \gamma_{22}FPE_j + \gamma_{23}GNP_j{*}FPE_j + \mu_{2j}$$

Where:

GNP_j = gross national product for country j

FPE_j = effectiveness of family planning programming for country j

GNP_j*FPE_j = interaction between GNP and FPE for country j

γ_{00} = the expected value of β_{0j} for women from countries with a GNP and FPE equal to 0

γ_{01} = the expected change in β_{0j} per unit change in GNP

γ_{02} = the expected change in β_{0j} per unit change in FPE

γ_{03} = the expected change in β_{0j} per unit change in GNP*FPE

γ_{10} = the expected value of β_{1j} for women from countries with a GNP and FPE equal to 0

γ_{11} = the expected change in β_{1j} per unit change in GNP

γ_{12} = the expected change in β_{1j} per unit change in FPE

γ_{13} = the expected change in β_{1j} per unit change in GNP*FPE

γ_{20} = the expected value of β_{2j} for women from countries with a GNP and FPE equal to 0

γ_{21} = the expected change in β_{2j} per unit change in GNP

γ_{22} = the expected change in β_{2j} per unit change in FPE

γ_{23} = the expected change in β_{2j} per unit change in GNP*FPE

Substituting the country-level equations into the individual-level equations, the combined model becomes:

$$\text{Logit}(EVER_{ij}) = \gamma_{00} + \gamma_{01}GNP_j + \gamma_{02}FPE_j + \gamma_{03}GNP_j{*}FPE_j$$
$$+ (\gamma_{10} + \gamma_{11}GNP_j + \gamma_{12}FPE_j + \gamma_{13}GNP_j{*}FPE_j)WED_{ij}$$

$$+ \ (\gamma_{20} + \gamma_{21}GNP_j + \gamma_{22}FPE_j + \gamma_{23}GNP_j{*}FPE_j)URBC_{ij}$$

$$+ \ \mu_{0j} + \mu_{1j}WED_{ij} + \mu_{2j}URBC_{ij}$$

Using these models, Wong and Mason obtained EB estimates of the logistic regression coefficients for each country, and their study's results suggested that these estimates were more interpretable than the classical estimates. In addition to main effects at the individual level, Wong and Mason found a strong GNP effect and cross-level interactions between childhood residence and wife's education and both GNP and FPE. For example, they found that even though women with more education were more likely than those with less education to practice contraception, countries with effective family planning programs offset this education difference. This information resulted in smaller country-level coefficients that represented the expected change in the log-odds of using contraception for every unit change in wife's education (β_{1j}) for countries with effective family planning programs, providing more accurate estimation of the probability of practicing contraception.

Wong and Mason, however, failed to center their measures, leading to intercept values that are difficult to interpret. For example, in their model, γ_{00} represents the expected value of the log-odds of using contraception for rural individuals with no education from countries with a gross national product of zero and a family planning effectiveness score of zero. While it is not clear how family planning was measured, it is difficult to imagine a country with a gross national product of zero; therefore, this intercept is meaningless. It would have been conceptually more helpful to use a centered gross national product score, so that the intercept reflected those countries with an average gross national product (Aiken & West, 1991; Burton, 1993, Kreft & de Leeuw, 1998; Kreft, de Leeuw, & Aiken, 1995).

G. Preliminary Study

A pilot study conducted prior to this dissertation indicated that some of the variability in the slopes and intercepts obtained from an approximation of a hierarchical logistic regression model used to predict NCLEX-RN performance from DRT scores could be explained by a school-level variable, type of nursing program (Schmidt, 1997). In an examination of a national sample of 5,698 nursing licensure candidates in 135 nursing schools, an approximation of a hierarchical logistic regression model was obtained by computing the logistic regression slopes and intercepts separately for each school, and then using these slopes and intercepts as dependent variables in two separate multiple regression equations, using second-level predictors as independent variables. Results indicated that the school-level variable, program type, accounts for approximately 9% of the variability in the intercepts and approximately 7% of the variability in the slopes. It was noted as a limitation to the approximation method, however, that these are estimates of the total school-level variance. If the error variance at this level is large, the school-level variable, program type, may be accounting for a great deal of true variance - but with this approximation method, there is no way of knowing.

It was hypothesized that the application of a full hierarchical logistic model would result in an improved analysis. For example, the estimation procedure used in hierarchical logistic regression models should result in more accurate coefficient estimates for each school. Treating each school separately, as was done in the pilot study, can result in small sample sizes, which makes ordinary least squares estimates very unstable, particularly from administration to administration (see Rubin, 1980 for a full discussion of the "bouncing beta" problem). In the context of linear models, EB estimates have generally been shown to be superior to ordinary least squares estimates, and generally result in more stable estimates for schools with smaller samples.

In addition, since using a hierarchical logistic regression model allows for a separation of error variance from true variance at the school level, it was hypothesized that a much clearer picture of the effect of type of program would emerge. Consequently, specific policy implications and recommendations could be developed based on sound data, and applicable to a wide variety of nursing programs.

CHAPTER III: METHOD

A. The Problem

The review of the literature on predicting performance on the NCLEX-RN revealed that other than the Akers (1992) study, there is nothing available other than the user's guide published by the National League for Nursing (NLN, 1989) that addresses the predictive validity of the Diagnostic Readiness Test. Akers (1992) found that the DRT is a significant predictor for both associate degree students and baccalaureate students in Mississippi, and indeed, the NLN has documented the predictive validity of its examination using a national sample of students across all three types of nursing programs (NLN, 1989). However, the DRT is not the only examination available that can assess cumulative performance in nursing school and provide an estimate of a candidate's probability of passing the NCLEX-RN (for example, see Beitz, 1994 for a comparison of the DRT to Mosby's Assess Test (Saxton, Pelikan, Nugent, & Needleman, 1989)). In deciding which assessment instrument to purchase, schools must maximize return on their investment. That is, they should purchase the examination that is, among other criteria, the most valid predictor of the NCLEX-RN for their population. Published research suggests that there is school-level variability as well as individual-level variability in predicting performance on the NCLEX-RN (Foti & De Yong, 1991), and the pilot study conducted prior to this study indicated that there is differential predictive validity for the DRT by different types of nursing programs (Schmidt, 1997). Therefore, schools need information not only on the overall predictive validity of the DRT for all students, but also on the predictive validity of the DRT for their particular type of institution in order to make an informed decision about which examination to invest in. Let us now turn our attention to the way in which the predictive validity of the DRT has traditionally been established.

Experimental data is collected on the DRT every two years during its development phase. In addition to providing the data necessary to calculate the probability of passing the NCLEX-RN at each DRT score point, the experimental data is used to establish the predictive validity of the DRT. Traditionally, a logistic regression analysis had been performed to obtain the validity coefficient for the DRT. The overall classification rate (success of the model in predicting actual pass/fail status) was reported, as well as a table of DRT scores and corresponding probabilities of passing the NCLEX-RN, which were calculated based on the results of the logistic regression. In addition, individual schools could request "customized" statistics based only on that school's population. As the number of schools requesting this service increased, it became clear that there was a large variation between schools in the logistic coefficient for DRT scores predicting pass/fail on the NCLEX-RN, and consequently a large variation in the statistics that were reported. In addition, the NCLEX-RN pass/fail data obtained by the NLN supported previous data suggesting that the pass rate on the licensure examination varied as a function of type of nursing program.

Both this high variability in logistic regression coefficients across schools and the apparent importance of a school-level variable, type of nursing program, in predicting pass/fail status on the NCLEX-RN suggests that the relationship between DRT and NCLEX-RN performance can best be modeled using a two-level hierarchical logistic regression approach. The primary school-level variable that was employed in this study is type of nursing program. An additional individual-level variable, NLN Pre-Admissions score, was included in some of the models to explore how the relationship between the DRT and NCLEX-RN may vary after controlling for initial student ability. An example of a hypothesis that could be tested by including NLN Pre-Admissions scores is that baccalaureate programs may have lower passing rates on the NCLEX-RN because they enroll students with lower initial ability. Although this may seem counter-intuitive, as the baccalaureate degree is the more prestigious degree, there has been more interest from potential nursing students in associate degree and diploma programs rather than baccalaureate programs in recent years. These programs require fewer years of course work before a candidate can become licensed to practice. In fact, associate degree programs have longer waiting lists than baccalaureate programs (NLN Press, 1995), indicating that they may very well be able to attract and admit a higher caliber of student.

B. Measures

The National League for Nursing's (NLN) Division of Assessment and Evaluation provides a variety of testing services to schools of nursing and to the greater nursing community. The Diagnostic Readiness Test (DRT) was designed to assess a nursing student's readiness to sit for the NCLEX-RN. The test consists of approximately 175 items (the number of items varies with different versions of the test). In addition to an analysis of individual strengths and weaknesses in specific content areas, each student is provided with a score that indicates their probability of passing the NCLEX-RN given their score on the DRT. The test is updated on a regular basis (i.e., every two years). All students who sit for the experimental version of the DRT and agree to release their pass/fail status on the NCLEX-RN to the NLN are included in the calculation of the probability score. Pass/fail status is obtained directly from each jurisdiction's licensing board.

The NLN also administers a Pre-Admission test (PreAd) to nursing school applicants. This test, consisting of 140 questions, measures verbal, mathematical, and general science ability. A standardized composite score is reported to the student as well as the school, and each school sets its own minimum passing score. The mean standardized composite score is 100, with a standard deviation of 10.

C. Data Collection

The total number of nursing students who sat for the DRT during the spring semester of 1993 and subsequently sat for the NCLEX-RN the following July was 11,570 students in 288 schools. However, in order to be able to control for initial ability, only students who had scores on the Pre-Admissions (PreAd) test were included in the analyses. This reduced the sample to a total of 5,698 students in 135 nursing schools across the nation[1].

The average number of students in each school was 42.21, with a standard deviation of 17.47, ranging from a low of 25 to a high of 135. Fifty-nine percent of the students were enrolled in associate degree programs, 26% in baccalaureate degree programs, and only 15% in diploma programs, which is representative of the program breakdown of both the 228 schools that used the PreAd and the DRT (62%, 23% and 15% respectively), as well as all nursing schools (64%, 28% and 8% respectively).

In order to facilitate interpretation of the results, DRT and average PreAd scores were centered approximately about the grand mean by subtracting 125

[1] According to the Nursing Data Review 1995 (NLN Press, 1995) and other National League for Nursing publications (NLN 1984; 1988), there were a total of 228 nursing schools that used both the NLN Pre-Admissions Test and the Diagnostic Readiness Test in 1993. Therefore, the current sample of 135 schools represents 59% of the population of nursing schools. A substantial number of schools refused to release NCLEX-RN performance data on the grounds of confidentiality, even though permission to access this information was obtained from the students when they sat for the DRT. There was no systematic refusal by program type; i.e.,

from each DRT score and 100 from each PreAd score to create the variables DRTC (centered DRT score) and PreAdC (centered PreAd score). The centering was done in order to improve the interpretability of the intercept coefficients, both at the individual and school level (Aiken & West, 1991; Burton, 1993; Kreft & de Leeuw, 1998; Kreft, de Leeuw, & Aiken, 1995). For example, consider the following two logistic regression analyses. In the first analysis, DRT scores were used as predictors of NCLEX-RN performance, while in the second, centered DRT scores were used. Both analyses were conducted using the same sample of 5698 subjects. Results appear below.

For DRT scores:

Logit(NCLEX-RN) = -3.9652 + .0482(DRT)

For centered DRT scores:

Logit(NCLEX-RN) = 2.0568 + .0482(DRTC)

The coefficients for the slope are the same for both equations, as would be expected. However, the coefficients for the intercepts are quite different, and need to be interpreted differently. For the first equation, the intercept coefficient represents the Logit(NCLEX-RN) for students with a DRT score of zero. According to this equation, students with a score of zero have only a 1.8-% chance of passing the NCLEX-RN. However, a score of zero is almost

---

one type of nursing program was not any more likely to refuse to provide data than any other impossible to attain on the DRT. Attempting to explain the variance across schools for such a meaningless coefficient is not a useful exercise.

In the second equation the intercept coefficient has a very different meaning. This coefficient represents the log odds of passing the NCLEX-RN for students with average DRT scores. In other words, a student with a DRT score of 125 has an 88.7% chance of passing the NCLEX-RN. Centering about the grand mean is clearly preferable, and creates intercept coefficients that can be meaningfully interpreted. In addition to this benefit, some researchers have suggested that centering about the grand mean results in improved explanation of between-school variance in hierarchical linear models (Schumacker & Bembry, 1995). Exploring the variability between schools in these intercept coefficients also becomes a more interesting task: what characteristics of a school might be expected to influence the chance of two students with average DRT scores passing the NCLEX-RN? This is the question that will be addressed in the second-level analyses.

D. The Models

Several hierarchical logistic regression models were specified for the analysis. The models were analyzed and parameter estimates obtained using the Mln software package (Rasbash & Woodhouse, 1996). This program allows

---

type.

for different parameter estimation methods with logistic regression models. Parameter estimates using only the fixed part of the model (MQL) were compared to those estimated from the fixed part plus higher level residuals (PQL), and first and second order approximations were compared as well. Previous research indicates that second order PQL estimates are more statistically efficient and less biased than the MQL and first order approximations (Goldstein & Rasbash, 1995; Paterson, 1995). However, it has been suggested that, since second order PQL estimates don't always converge, an appropriate strategy is to model the simpler estimation procedures (i.e., first order MQL) and switch to more complex procedures as the model becomes established (Paterson, 1995).

Models were specified from the simplest to the most complex and were compared in order to determine the model that best fit the data using a variety of methods. The significance of each of the fixed parameters was determined by dividing the parameter estimate by its standard error. However, the distribution of the ratio of the random parameters to their standard errors may depart from normal, so it has been suggested that the likelihood ratio statistic be used to compare models instead (Woodhouse, Rasbash, Goldstein, & Yang, 1995). However, Goldstein and Rasbash (1995) have suggested that the likelihoods should not be used to make comparisons between discrete response models because the computation of the likelihood relies on the linearization approximation to the log-likelihood involved in the model estimation, and therefore comparisons may be unreliable. As will be seen later in the results section, the likelihoods did not follow an expected pattern, and the final model was determined by examining the fullest model possible and eliminating non-significant terms. In addition to these procedures, residual plots were examined visually. In the following sections, we describe each family of models that were analyzed.

Model 1

Beginning with the simplest possible model, a model analogous to a one-way ANOVA with random effects, the overall probability that a student will pass the NCLEX-RN can be modeled as follows at the first level:

$\text{Logit}(NCLEX_{ij}) = \beta_{0j}$

Where:

$\text{Logit}(NCLEX_{ij})$ = log-odds of passing NCLEX for student i in school j [the log-odds is defined as the natural logarithm of the probability of passing NCLEX divided by 1-(the probability of passing NCLEX); $\text{Logit}(NCLEX_{ij}) = \ln\{p(NCLEX_{ij})/[1-p(NCLEX_{ij})]\}$

β_{0j} = intercept term for school j, or the expected value of the log-odds of passing NCLEX for any individual

A more general model that allows the intercept to vary across schools can be developed:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

Where:

γ_{00} = the expected value of β_{0j}

Substituting the school-level equations into the student-level equations, the combined model becomes:

$$\text{Logit}(\text{NCLEX}_{ij}) = \gamma_{00} + \mu_{0j}$$

This model, known as the fully unconditional model (Bryk & Raudenbush, 1992) provides the baseline prediction of NCLEX-RN performance as well as a baseline value for the likelihood that can be compared with all subsequent models.

Models 2.1-2.3

In this second family of models, analogous to a one-factor ANCOVA with random effects, a single level-one predictor, centered DRT scores, is included. No level-two predictors are included. At the first level, the fullest of these models can be specified as follows:

$$\text{Logit}(\text{NCLEX}_{ij}) = \beta_{0j} + \beta_{1j}\text{DRTC}_{ij}$$

Where:

β_{0j} = intercept term for school j, or the expected value of the log-odds of passing NCLEX for an individual for whom the value of DRTC is zero

β_{1j} = slope for DRTC score in school j, or the expected change in the log-odds of passing NCLEX for every unit change in DRTC

$DRTC_{ij}$ = DRTC score for student i in school j

A more general model that allows the intercepts to vary across schools can be developed:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \mu_{1j}$$

Where:

γ_{00} = the expected value of β_{0j}

γ_{10} = the expected value of β_{1j}

Substituting the school-level equations into the student-level equations, the combined model becomes:

$$\text{Logit}(NCLEX_{ij}) = \gamma_{00} + \mu_{0j} + \gamma_{10}\, DRTC_{ij} + \mu_{1j} DRTC_{ij}$$

In this model (2.3), both the slopes and intercepts are allowed to vary. In simpler forms of this model, such as in model 2.1, only the intercepts are allowed to vary, while in model 2.2, only the slopes are allowed to vary.

Models 3.1-3.3

The third family of models is similar to the second, except that the single level-one predictor included is centered PreAd score rather than centered DRT scores. No level-two predictors are included. At the first level, the fullest of these models can be specified as follows:

$$\text{Logit}(\text{NCLEX}_{ij}) = \beta_{0j} + \beta_{2j}\text{PreAdC}_{ij}$$

Where:

β_{0j} = intercept term for school j, or the expected value of the log-odds of passing NCLEX for an individual for whom the value of PreAdC is zero

β_{2j} = slope for PreAdC score in school j, or the expected change in the log-odds of passing NCLEX for every unit change in PreAdC

$PreAdC_{ij}$ = PreAdC score for student i in school j

A more general model that allows the intercepts and slopes to vary across schools can be developed:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\beta_{2j} = \gamma_{20} + \mu_{2j}$$

Where:

γ_{00} = the expected value of β_{0j}

γ_{20} = the expected value of β_{2j}

Substituting the school-level equations into the student-level equations, the combined model becomes:

$$\text{Logit}(NCLEX_{ij}) = \gamma_{00} + \mu_{0j} + \gamma_{20}\,PreAdC_{ij} + \mu_{2j}PreAdC_{ij}$$

Again, in this model (3.3), both the slopes and intercepts are allowed to vary. In simpler forms of this model, such as in model 3.1, only the intercepts are allowed to vary, while in model 3.2, only the slopes are allowed to vary.

Models 4.1-4.7

The fourth family of models combines the second and third families, and includes both level-one predictors: centered DRT and centered PreAd scores. No level-two predictors are included. At the first level, the fullest of these models can be specified as follows:

$$\text{Logit}(\text{NCLEX}_{ij}) = \beta_{0j} + \beta_{1j}\text{DRTC}_{ij} + \beta_{2j}\text{PreAdC}_{ij}$$

Where:

β_{0j}	=	intercept term for school j, or the expected value of the log-odds of passing NCLEX for an individual for whom the value of PreAdC and DRTC is zero
β_{1j}	=	slope for DRTC score in school j, or the expected change in the log-odds of passing NCLEX for every unit change in DRTC
DRTC_{ij}	=	DRTC score for student i in school j
β_{2j}	=	slope for PreAdC score in school j, or the expected change in the log-odds of passing NCLEX for every unit change in PreAdC
PreAdC_{ij}	=	PreAdC score for student i in school j

A more general model that allows the intercepts and slopes to vary across schools can be developed:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \mu_{1j}$$

$$\beta_{2j} = \gamma_{20} + \mu_{2j}$$

Where:

γ_{00} = the expected value of β_{0j}

γ_{10} = the expected value of β_{1j}

γ_{20} = the expected value of β_{2j}

Substituting the school-level equations into the student-level equations, the combined model becomes:

$$\text{Logit}(\text{NCLEX}_{ij}) = \gamma_{00} + \mu_{0j} + \gamma_{10}\,\text{DRTC}_{ij} + \mu_{1j}\text{DRTC}_{ij} + \gamma_{20}\text{PreAdC}_{ij} + \mu_{2j}\text{PreAdC}_{ij}$$

In this model (4.7), both the slopes for DRTC and PreAdC, as well as the intercepts are allowed to vary. In simpler forms of this model, such as in model 4.1, only the intercepts are allowed to vary. In model 4.2, only the DRTC slopes are allowed to vary, while in model 4.3, the DRTC slopes are allowed to vary along with the intercepts. In model 4.4, only the PreAdC intercepts are allowed to vary, while in model 4.5 the PreAdC slopes are allowed to vary along with the intercepts. In model 4.6, both the DRTC and PreAdC slopes are allowed to vary.

<u>Model 5</u>

Model 5 is the simplest possible model that contains a level-two predictor, and is analogous to the means as outcomes regression model (Bryk & Raudenbush, 1992). At the first level, the overall probability that a student will pass the NCLEX-RN can be modeled as follows:

$$\text{Logit}(\text{NCLEX}_{ij}) = \beta_{0j}$$

Where:

β_{0j} = intercept term for school j, or the expected value of the log-odds of passing NCLEX for any individual

A more general model that allows the intercepts to vary across schools and includes type of nursing program as a possible predictor of this variability can be developed:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j + \mu_{0j}$$

Where:

Program1_j = an indicator variable to code program type; 1 if program type is diploma, zero otherwise

Program2_j = an indicator variable to code program type; 1 if program type is associate degree, zero otherwise

γ_{00} = the expected value of β_{0j} for Baccalaureate students (Program1 = 0 and Program2 = 0; see coding above)

γ_{01} = the expected difference in β_{0j} between baccalaureate students and diploma students

γ_{02} = the expected difference in β_{0j} between baccalaureate students and associate degree students

Substituting the school-level equations into the student-level equations, the combined model becomes:

$$\text{Logit}(\text{NCLEX}_{ij}) = \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j + \mu_{0j}$$

This model provides information on the degree to which type of nursing program can predict the variability in intercepts from the fully unconditional model. (Note: as will be seen later in the results section, it was possible to collapse type of nursing program into two categories: baccalaureate and non-baccalaureate.)

Models 6.1-6.9

In this family of models a single level-one predictor, centered DRT scores, is included along with the level-two predictor, type of nursing program. At the first level, the fullest of these models can be specified as follows:

$$\text{Logit}(\text{NCLEX}_{ij}) = \beta_{0j} + \beta_{1j}\text{DRTC}_{ij}$$

Where:

β_{0j} = intercept term for school j, or the expected value of the log-odds of passing NCLEX for an individual for whom the value of DRTC is zero

β_{1j} = slope for DRTC score in school j, or the expected change in the log-odds of passing NCLEX for every unit change in DRTC

$DRTC_{ij}$ = DRTC score for student i in school j

A more general model that incorporates the school-level variable 'type of nursing program' can be developed:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j + \mu_{0j},$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\text{Program1}_j + \gamma_{12}\text{Program2}_j + \mu_{1j}$$

Where:

Program1_j and Program2_j are defined as before

γ_{00} = the expected value of β_{0j} for Baccalaureate students (Program1 = 0 and Program2 = 0)

γ_{01} = the expected difference in β_{0j} between baccalaureate students and diploma students

γ_{02} = the expected difference in β_{0j} between baccalaureate students and associate degree students

γ_{10} = the expected value of β_{1j} for Baccalaureate students (Program1 = 0 and Program2 = 0; see coding above).

γ_{11} = the expected difference in β_{1j} between baccalaureate students and diploma students

γ_{12} = the expected difference in β_{1j} between baccalaureate students and associate degree students

Substituting the school-level equations into the student-level equations, the combined model becomes:

$$\begin{aligned}\text{Logit}(\text{NCLEX}_{ij}) &= \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j \\ &+ (\gamma_{10} + \gamma_{11}\text{Program1} + \gamma_{12}\text{Program2})\text{DRTC}_{ij} \\ &+ \mu_{0j} + \mu_{1j}\text{DRTC}_{ij}\end{aligned}$$

Multiplying out the terms, the combined general model becomes:

$$\begin{aligned}\text{Logit}(\text{NCLEX}_{ij}) &= \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j \\ &+ \gamma_{10}\text{DRTC}_{ij} + \gamma_{11}\text{Program1}_j\text{DRTC}_{ij} + \gamma_{12}\text{Program2}_j\text{DRTC}_{ij} \\ &+ \mu_{0j} + \mu_{1j}\text{DRTC}_{ij}\end{aligned}$$

In this model (6.9), both the slopes and intercepts are allowed to vary, and type of nursing program predicts both the β_{0j}'s and the β_{1j}'s. In simpler forms of this model, such as in model 6.1, only the intercepts are allowed to vary and type of nursing program predicts the β_{0j}'s only. In model 6.2, only the slopes are allowed to vary while type of nursing program predicts the β_{0j}'s only. In model 6.3, both the slopes and intercepts are allowed to vary, and type of nursing program predicts the β_{0j}'s only. Models 6.4 through 6.6 follow the same pattern, except that type of nursing program predicts the β_{1j}'s only. Finally, in models 6.7 through 6.9, the intercepts and slopes follow the same pattern, and type of nursing program predicts both the β_{0j}'s and the β_{1j}'s.

Models 7.1-7.9

The seventh family of models is similar to the sixth, except that the single level-one predictor included is centered PreAd score rather than centered DRT scores along with the level-two predictor, type of nursing program. At the first level, the fullest of these models can be specified as follows:

$$\text{Logit}(\text{NCLEX}_{ij}) = \beta_{0j} + \beta_{2j}\text{PreAdC}_{ij}$$

Where:

β_{0j} = intercept term for school j, or the expected value of the log-odds of passing NCLEX for an individual for whom the value of PreAdC is zero

β_{2j} = slope for PreAdC score in school j, or the expected change in the log-odds of passing NCLEX for every unit change in PreAdC

PreAdC_{ij} = PreAdC score for student i in school j

A more general model that incorporates the school-level variable 'type of nursing program' can be developed:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j + \mu_{0j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\text{Program1}_j + \gamma_{22}\text{Program2}_j + \mu_{2j}$$

Where:

$Program1_j$ and $Program2_j$ are defined as before

γ_{00} = the expected value of β_{0j} for Baccalaureate students (Program1 = 0 and Program2 = 0)

γ_{01} = the expected difference in β_{0j} between baccalaureate students and diploma students

γ_{02} = the expected difference in β_{0j} between baccalaureate students and associate degree students

γ_{20} = the expected value of β_{2j} for Baccalaureate students (Program1 = 0 and Program2 = 0; see coding above).

γ_{21} = the expected difference in β_{2j} between baccalaureate students and diploma students

γ_{22} = the expected difference in β_{2j} between baccalaureate students and associate degree students

Substituting the school-level equations into the student-level equations, the combined model becomes:

$$\text{Logit}(\text{NCLEX}_{ij}) = \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j$$

$$+ \quad (\gamma_{20} + \gamma_{21}\text{Program1}_j + \gamma_{22}\text{Program2}_j)\text{PreAdC}_{ij}$$

$$+ \quad \mu_{0j} + \mu_{2j}\text{DRTC}_{ij}$$

Multiplying out the terms, the combined general model becomes:

$$Logit(NCLEX_{ij}) = \gamma_{00} + \gamma_{01}Program1_j + \gamma_{02}Program2_j$$

$$+ \gamma_{20}PreAdC_{ij} + \gamma_{21}Program1_j PreAdC_{ij} + \gamma_{22}Program2_j PreAdC_{ij}$$

$$+ \mu_{0j} + \mu_{2j}PreAdC_{ij}$$

In this model (7.9), both the slopes and intercepts are allowed to vary, and type of nursing program predicts both the β_{0j}'s and the β_{1j}'s. In simpler forms of this model, such as in model 7.1, only the intercepts are allowed to vary and type of nursing program predicts the β_{0j}'s only. In model 7.2, only the slopes are allowed to vary while type of nursing program predicts the β_{0j}'s only. In model 7.3, both the slopes and intercepts are allowed to vary, and type of nursing program predicts the β_{0j}'s only. Models 7.4 through 7.6 follow the same pattern, except that type of nursing program predicts the β_{1j}'s only. Finally, in models 7.7 through 7.9, the intercepts and slopes follow the same pattern, and type of nursing program predicts both the β_{0j}'s and the β_{1j}'s.

Models 8.1-8.54

The eighth and final family of models combines the sixth and seventh families, and include both level-one predictors (centered DRT and centered PreAd scores) along with the second-level predictor, type of nursing program. At the first level, the fullest of these models can be specified as follows:

$$Logit(NCLEX_{ij}) = \beta_{0j} + \beta_{1j}DRTC_{ij} + \beta_{2j}PreAdC_{ij}$$

Where:

β_{0j} = intercept term for school j, or the expected value of the log-odds of passing NCLEX for an individual for whom the value of DRTC and PreAdC is zero

β_{1j} = slope for DRTC score in school j, or the expected change in the log-odds of passing NCLEX for every unit change in DRTC

$DRTC_{ij}$ = DRTC score for student i in school j

β_{2j} = slope for PreAdC score in school j, or the expected change in the log-odds of passing NCLEX for every unit change in PreAdC

$PreAdC_{ij}$ = PreAdC score for student i in school j

A more general model that incorporates the school-level variable 'type of nursing program' can be developed:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\text{Program1}_j + \gamma_{12}\text{Program2}_j + \mu_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\text{Program1}_j + \gamma_{22}\text{Program2}_j + \mu_{2j}$$

Where:

Program1_j and Program2_j are defined as before

γ_{00} = the expected value of β_{0j} for Baccalaureate students (Program1 = 0 and Program2 = 0)

γ_{01} = the expected difference in β_{0j} between baccalaureate students and diploma students

γ_{02} = the expected difference in β_{0j} between baccalaureate students and associate degree students

γ_{10} = the expected value of β_{1j} for Baccalaureate students (Program1 = 0 and Program2 = 0; see coding above).

γ_{11} = the expected difference in β_{1j} between baccalaureate students and diploma students

γ_{12} = the expected difference in β_{1j} between baccalaureate students and associate degree students

γ_{20} = the expected value of β_{2j} for Baccalaureate students (Program1 = 0 and Program2 = 0; see coding above).

γ_{21} = the expected difference in β_{2j} between baccalaureate students and diploma students

γ_{22} = the expected difference in β_{2j} between baccalaureate students and associate degree students

Substituting the school-level equations into the student-level equations, the combined model becomes:

$$\text{Logit}(\text{NCLEX}_{ij}) = \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j$$
$$+ \quad (\gamma_{10} + \gamma_{11}\text{Program1}_j + \gamma_{12}\text{Program2}_j)\text{DRTC}_{ij}$$
$$+ \quad (\gamma_{20} + \gamma_{21}\text{Program1}_j + \gamma_{22}\text{Program2}_j)\text{PreAdC}_{ij}$$
$$+ \quad \mu_{0j} + \mu_{1j}\text{DRTC}_{ij} + \mu_{2j}\text{PreAdC}_{ij}$$

Multiplying out the terms, the combined general model becomes:

$$\text{Logit}(\text{NCLEX}_{ij}) = \gamma_{00} + \gamma_{01}\text{Program1}_j + \gamma_{02}\text{Program2}_j$$
$$+ \gamma_{10}\,\text{DRTC}_{ij} + \gamma_{11}\text{Program1}_j\,\text{DRTC}_{ij} + \gamma_{12}\text{Program2}_j\,\text{DRTC}_{ij}$$
$$+ \gamma_{20}\text{PreAdC}_{ij} + \gamma_{21}\text{Program1}_j\,\text{PreAdC}_{ij} + \gamma_{22}\text{Program2PreAdC}_{ij}$$
$$+ \mu_{0j} + \mu_{1j}\text{DRTC}_{ij} + \mu_{2j}\text{PreAdC}_{ij}$$

In this model (8.54), both the slopes for DRTC and PreAdC, as well as the intercepts are allowed to vary. The number of sub-models are too numerous to outline in detail, but contain all possible combinations of variation in slopes and intercepts, with type of nursing program predicting various combinations of β_{0j}'s, β_{1j}'s and β_{2j}'s.

F. <u>Hypotheses and Model Interpretation</u>

It should be noted that if the parameters associated with the variation in the school-level effects in the model (e.g., μ_{1j}) are set to zero, we are left with the typical classical logistic regression equation predicting NCLEX-RN performance from DRTC and PreAdC scores across schools. It is hypothesized, however, that there will be significant school-level variability in both the slopes and the intercepts that can be partially explained by type of nursing program. In such a model, the parameters associated with the relationship between DRTC and NCLEX-RN (i.e., γ_{10}, γ_{11}, and γ_{12}) can be viewed as the differential validity coefficients for the DRTC by type of nursing program after controlling for initial student ability, PreAdC. If any of these parameters, γ_{10}, γ_{11}, or γ_{12}, are significantly different from zero, it will indicate that the validity of the DRTC as a predictor of NCLEX-RN performance varies as a function of program type after controlling for initial student ability, and validity coefficients should therefore be reported separately for each of the nursing programs. However, it may be the case that after initial student ability is taken into account, the parameters associated with program type may no longer be significant, indicating that the DRT's relationship to the NCLEX is not affected as much by the type of nursing program, but by the ability of the student. It is hypothesized, however, that program type will account for a significant proportion of the school-level variance in the ability of the DRT to predict NCLEX-RN performance, even after initial student ability is taken into account.

With the final model obtained, the following procedures were conducted to facilitate interpretation of the results. First, for each of the schools, EB estimates of the parameters were compared to the classical logistic regression coefficients. In addition, to provide a sense of the predictive validity of an examination with a dichotomous criterion, the probability of passing the NCLEX-RN at different percentile ranks of DRT score was calculated and compared. If the DRT is a significant predictor of NCLEX-RN performance, it should be able to discriminate between those who pass and those who fail the NCLEX-RN, and so the probability of passing the NCLEX-RN should vary at different percentiles for each of the programs.

CHAPTER IV: RESULTS

A. Initial Data Exploration

Table 1 presents the descriptive statistics on the three measures of interest: DRT scores, Pre-Admission scores, and NCLEX-RN results.

Table 1

Descriptive Statistics
(n=5698)

Variable	Mean	SD	Minimum	Maximum
DRT Score	124.2	14.15	36	172
Pre-Admission Score	99.3	8.11	68	141

NCLEX Results	Frequency	Percent
Pass	4929	86%
Fail	769	14%

Table 2 presents the results of a Chi-square analysis performed to determine the extent to which pass/fail status on the NCLEX-RN varies as a function of type of nursing program.

Table 2

Pass/Fail Status by Type of Nursing Program

	Diploma Programs	Associate Degree Programs	Baccalaureate Programs	χ^2,(df),p
Pass	87%	88%	83%	15.09,(2), .00053

While it may not seem as if these results have much practical significance, it must be remembered that nursing schools consider these differences quite significant. (Indeed, there are jurisdictions that cut off funding to schools when their pass rate falls below a certain percentage.) Prior to running the hierarchical logistic regression models, a series of classical logistic regression models were specified to predict pass/fail status from centered DRT scores alone, centered Pre-Admission scores alone, and centered DRT scores together with centered Pre-Admission scores, for the entire sample. Results of these analyses can be found in Table 3. As can be seen from these results, in an equation that contains both centered DRT and centered PreAd scores, centered DRT scores are the only significant predictor of pass/fail status. These results seem to clearly support the predictive validity of the DRT.

Table 3

Results of Logistic Regression Analyses for the Overall Sample (n=5698)

Model	B	S.E.	Wald	Df	Sig	R	Exp(B)
Logit(NCLEX$_{ij}$) = $\beta_0 + \beta_1$DRTC$_i$							
β_0	2.06	.005	2097.3	1	<.000		
β_1	.05	.003	275.0	1	<.000	.246	1.05
Logit(NCLEX$_{ij}$) = $\beta_0 + \beta_1$PreAdC$_i$							
β_0	1.89	.04	2251.8	1	<.000		
β_1	.02	.005	15.8	1	<.000	.005	1.02
Logit(NCLEX$_{ij}$) = $\beta_0 + \beta_1$DRTC$_i$ + β_2PreAdC$_i$							
β_0	2.06	.045	2082.5	1	<.000		
β_1	.05	.003	273.3	1	<.000	.241	1.05
β_2	.007	.005	1.7	1	.196	.000	1.007

Classical logistic regression equations were also specified for each school so that the coefficients obtained could be compared to the EB estimates that were obtained in subsequent hierarchical models. Coefficients for logistic regression equations predicting performance on the NCLEX-RN from centered DRT scores were estimated separately for each of the 135 schools in the sample. Pre-Admission scores were not included in this analysis because they did not appear to significantly improve prediction of NCLEX-RN performance (see Table 3). Each slope coefficient represents the rate of change in the log-odds for passing NCLEX-RN per unit change in DRT score. Each intercept coefficient is the log-odds of passing NCLEX-RN for a student with an average DRT score. As can be seen from Appendix A, there is a considerable amount of variability in the slopes and intercepts. It is notable that in some cases, the slope coefficient is negative, indicating that the better a student does on the DRT, the lower their chances are of passing the NCLEX-RN!

As part of a pilot study (Schmidt, 1997), an approximation of a hierarchical logistic regression model was specified by using these coefficients as dependent measures in two separate multiple regression equations. The independent variables that predicted variability in these coefficients were school-level variables such as type of nursing program (diploma, associate degree, or baccalaureate), enrollment, accreditation status, and whether the school was public or private. The results of these analyses appear in Table 4. Descriptive statistics for all level-two variables appear in Table 5.

Table 4

Results of Multiple Regressions

	Predicting Intercept Coefficients	p	Predicting Slope Coefficients	p
Constant	2.857	<.000	.0522	.102
Program Dummy 1 (1=diploma, 0 otherwise)	-.9933	.0024	-.0731	.0057
Program Dummy 2 (1=associate degree, 0 otherwise)	-.7110	.0022	-.0447	.0164
Enrollment	.0000	.9716	.0000	.3897
Public or Private? (1=private, 0=public)	.4199	.0943	.0196	.3322
Accredited or Not? (1=accred, 0=no)	.1733	.5755	.0146	.5598
R, R^2	.34151, .1166		.29803, .0888	

Table 5

Descriptive Statistics on School Level Variables

	Frequency	Percentage
Type of Program:		
Diploma	18	13%
Associate Degree	84	62%
Baccalaureate	33	25%
Public or Private?		
Public	51	38%
Private	84	62%
Accredited or not?		
Yes	120	89%
No	15	11%
Enrollment	mean: 250.44 sd: 163.14 range: 45-1035 skew: 2.495 kurt: 8.749	

An examination of the Table 4 results suggests that, in the context of the predictors included, the only significant predictor of variability in the slopes and intercepts is the type of nursing program. In order to develop meaningful coefficients that could predict first-level slopes from type of nursing program, analyses were conducted to examine whether the three types of nursing programs could be collapsed into two categories. An examination of the results in Table 4 suggests that Baccalaureate programs are different from both Diploma and Associate Degree programs, but that Associate and Diploma programs are not that different from one another. For example, if the significant predictors of the first-level slope and intercept coefficients from Table 4 are used to create two regression equations (i.e., for intercepts: Y = 2.857 - .9933(dummy code for Diploma programs) - .7110(dummy code for Associate Degree programs), and for slopes: Y = .0522 - .0731(dummy code for Diploma programs) - .0447(dummy code for Associate Degree programs)), the predicted first-level intercept coefficient for each type of program would be: 2.857 for Baccalaureate programs; 1.86737 for Diploma programs; and 2.146 for Associate Degree programs, and the predicted first-level slope coefficient for each type of program would be .0522, -.0209, and .0075, respectively. To assess whether it was useful to distinguish between Associate Degree programs and Diploma programs, two sets of analyses were conducted. In the first, two dummy codes were created to compare Diploma programs to the other two programs, while in the second, two dummy codes were created to compare Associate Degree programs to the other two programs. Regression analyses were conducted with each set of codes to examine whether the coefficients representing the differences between the various program types were significant. Results of this analysis can be found in Table 6.

Table 6

Results of Multiple Regression Analysis Comparing Programs

	Predicting First Level Intercept Coefficients	p	Predicting First Level Slope Coefficients	p
Analysis One:				
Constant	2.236	<.000	.019	.339
Program Type 1 (1=Associate, 0=Other)	.106	.695	.0125	.569
Program Type 2 (1=Baccaluareate, 0=Other)	.851	.006	.0631	.012
Analysis Two:				
Constant	2.343	.001	.0315	.001
Program Type 1 (1=Diploma, 0=Other)	-.106	.695	-.0125	.569
Program Type 2 (1=Baccaluareate, 0=Other)	.745	.001	.0506	.004

As can be seen by an examination of these results, the coefficients that compare Associate Degree and Diploma programs to each other were not significant in predicting either slopes or intercepts, while the coefficients comparing Baccalaureate programs to both Associate Degree and Diploma programs were significant for both slopes and intercepts. Consequently, type of nursing program was recoded to reflect Baccalaureate vs. Non-Baccalaureate programs.

Finally, multiple regression analyses were performed to predict the variability in first-level slopes and intercepts from the recoded variable, type of nursing program. Results of this analysis appear in Table 7.

Table 7

Results of Multiple Regressions with Re-Coded Type of Nursing Program

	Predicting First Level Intercept Coefficients	**p**	**Predicting First Level Slope Coefficients**	**p**
Constant	2.3240	<.000	.0293	.001
Type of Nursing Program (1=Baccalaureate, 0=Non-Baccalaureate)	0.7638	.0004	0.0528	.0021
R, R^2	.303, .092		.263, .069	

The results of these analyses were incorporated into the subsequent hierarchical logistic regression models in that type of nursing program, dichotomized into baccalaureate vs. non-baccalaureate, was the only second-level predictor included in the models.

B. Results of the Hierarchical Logistic Regression Analyses

Each of the models were analyzed using the Mln software package (Rasbash & Woodhouse, 1996). Parameter estimates presented below are the second order PQL estimates. MQL estimates and first order PQL estimates were also obtained, but the second order PQL estimates consistently provided the smallest likelihood values, and research has demonstrated that second order PQL estimates are more statistically efficient and less biased estimates than the MQL and first order approximations (Goldstein & Rasbash, 1995; Paterson, 1995). Parameter estimates for each of the models outlined in the methods section (except for models 8.1 through 8.54) follows. As was pointed out previously, the number of sub-models in the eighth family of models are too numerous to outline in detail, but contain all possible combinations of variation in slopes and intercepts, with type of nursing program predicting various combinations of β_{0j}'s, β_{1j}'s and β_{2j}'s. This family of models will be discussed more fully in the section on the determination of the final model.

Model 1

Model one is the simplest possible model analogous to a one-way ANOVA with random effects. This model, also known as the fully unconditional model (Bryk & Raudenbush, 1992) provides the baseline prediction of NCLEX-

RN performance as well as a baseline value for the likelihood that can be compared to all subsequent models. Second-order PQL estimates for this model are found in Table 8.

Table 8

Parameter Estimates for Model 1

	Model 1	
Parameter	**estimate**	**S.E.**
γ_{00}	1.891	.04761
$\tau_{00}(\mathrm{var}(\mu_{0j}))$	.09157	.03616
Likelihood	3695.77	

Note. Model 1 is specified as: $\mathrm{Logit}(NCLEX_{ij}) = \gamma_{00} + \mu_{0j}$

Models 2.1 - 2.3

In the second family of models, analogous to a one-factor ANCOVA with random effects, a single level-one predictor, centered DRT scores, is included. No level-two predictors are included. In the fullest of these models (2.3), both the slopes and intercepts are allowed to vary. In simpler forms of this model, such as in model 2.1, only the intercepts are allowed to vary, while in model 2.2, only the slopes are allowed to vary. Second order PQL estimates for these models are found in Table 9.

Table 9

Parameter Estimates for Models 2.1-2.3

	Model 2.1		**Model 2.2**		**Model 2.3**	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	2.057	.0467	2.11	.04692	2.11	.04696
τ_{00} **(var(μ_{0j}))**	0	0			0	0
γ_{10}	.04818	.00308	.04209	.003897	.04206	.0039
τ_{11} **(var(μ_{1j}))**			.00066	.00021	.00066	.00021
Likelihood	3614.2		3201.79		3201.69	

Note. The fullest form of model 2 is specified as: $\text{Logit}(\text{NCLEX}_{ij}) = \gamma_{00} + \gamma_{10}\text{DRTC}_{ij} + \mu_{0j} + \mu_{1j}\text{DRTC}_{ij}$

Models 3.1 - 3.3

The third family of models is similar to the second, except that the single level-one predictor included is centered PreAd score rather than centered DRT scores. Again, in the fullest of these models (3.3), both the slopes and intercepts are allowed to vary. In simpler forms of this model, such as in model 3.1, only the intercepts are allowed to vary, while in model 3.2, only the slopes are allowed to vary. Second order PQL estimates for these models are found in Table 10.

Table 10

Parameter Estimates for Models 3.1-3.3

	Model 2.1		**Model 2.2**		**Model 2.3**	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	1.912	.01891	1.898	.03972	1.906	.04532
τ_{00} **(var(μ_{0j}))**	.08822	.9822			.05892	.03244
γ_{20}	.0189	.00485	.01911	.00532	.01276	.00495
τ_{20} **(var(μ_{2j}))**			.00066	.00532	.00008	.00037
Likelihood	3662.63		3772.74		3744.07	

Note. The fullest form of model 3 is specified as: Logit(NCLEX$_{ij}$) = γ_{00} + γ_{20}PreAdC$_{ij}$ + μ_{0j} + μ_{2j}PreAdC$_{ij}$

Models 4.1 - 4.7

The fourth family of models combines the second and third families, and include both level-one predictors: centered DRT and centered PreAd scores. No level-two predictors are included. In the fullest of these models (4.7), both the slopes for DRTC and PreAdC, as well as the intercepts are allowed to vary. In simpler forms of this model, such as in model 4.1, only the intercepts are allowed to vary. In model 4.2, only the DRTC slopes are allowed to vary, while in model 4.3, the DRTC slopes are allowed to vary along with the intercepts. In model 4.4, only the PreAdC intercepts are allowed to vary, while in model 4.5 the PreAdC slopes are allowed to vary along with the intercepts. In model 4.6, both the DRTC and PreAdC slopes are allowed to vary. Second order PQL estimates for these models are found in Tables 11A through 11C.

Table 11A

Parameter Estimates for Model 4.1-4.3

	Model 4.1		**Model 4.2**		**Model 4.3**	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	2.062	.04807	2.116	.04733	2.116	.04733
τ_{00} (var(μ_{0j}))	0	0			0	0
γ_{10}	.04763	.00312	.04148	.00393	.04148	.00393
τ_{10} (var(μ_{1j}))			.00066	.00021	.00066	.00021
γ_{20}	.00649	.00534	.00692	.00527	.0007	.00527
τ_{01} (cov(μ_{0j}/μ_{1j}))					0	0
Likelihood	3631.64		3214.14		3214.14	

Note. The fullest form of model 4 is specified as: $\text{Logit}(\text{NCLEX}_{ij}) = \gamma_{00} + \gamma_{10}\text{DRTC}_{ij} + \gamma_{20}\text{PreAdC}_{ij} + \mu_{0j} + \mu_{1j}\text{DRTC}_{ij} + \mu_{2j}\text{PreAdC}_{ij}$

Table 11B

Parameter Estimates for Model 4.4-4.5

	Model 4.4		**Model 4.5**	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	2.066	.048	2.062	.04799
τ_{00} **(var(μ_{0j}))**			0	0
γ_{10}	.04766	.003118	.04763	.00312
γ_{20}	.00633	.00542	.00649	.00533
τ_{20} **(var(μ_{2j}))**	.0001	.00041	0	0
τ_{02}**(cov(μ_{0j}/μ_{2j}))**			0	0
Likelihood	3604.94		3631.64	

Note. The fullest form of model 4 is specified as: Logit(NCLEX$_{ij}$) = $\gamma_{00} + \gamma_{10}$DRTC$_{ij}$ + γ_{20}PreAdC$_{ij}$ + μ_{0j} + μ_{1j}DRTC$_{ij}$ + μ_{2j}PreAdC$_{ij}$

Table 11C

Parameter Estimates for Model 4.6-4.7

	Model 4.6		Model 4.7	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	2.116	.0473	2.116	.0473
τ_{00} (var(μ_{0j}))			0	0
γ_{10}	.04148	.00393	.04148	.00393
τ_{11} (var(μ_{1j}))	.00066	.00021	.00066	.00021
γ_{20}	.00692	.00527	.00692	.00527
τ_{22} (var(μ_{2j}))	0	0	0	0
τ_{01}(cov(μ_{0j}/μ_{1j}))	0	0	0	0
τ_{02}(cov(μ_{0j}/μ_{2j}))	0	0	0	0
τ_{12}(cov(μ_{1j}/μ_{2j}))			0	0
Likelihood	3214.25		3214.25	

Note. The fullest form of model 4 is specified as: Logit(NCLEX$_{ij}$) = γ_{00} + γ_{10}DRTC$_{ij}$ + γ_{20}PreAdC$_{ij}$ + μ_{0j} + μ_{1j}DRTC$_{ij}$ + μ_{2j}PreAdC$_{ij}$

Model 5

Model 5 is the simplest possible model that contains a level-two predictor, and is analogous to the means as outcomes regression model (Bryk &

Raudenbush, 1992). Second order PQL estimates for this model are found in Table 12.

Table 12

Parameter Estimates for Model 5

	Model 5	
Parameter	**est.**	**S.E.**
γ_{00}	1.965	.05403
τ_{00} (var(μ_{0j}))	.06641	.03282
γ_{01}	-.2941	.1001
Likelihood	3715.42	

Note. Model 5 is specified as: $\text{Logit}(NCLEX_{ij}) = \gamma_{00} + \gamma_{01}BACC_{ij} + \mu_{0j}$

Models 6.1 - 6.9

In this family of models a single level-one predictor, centered DRT scores, is included along with the level-two predictor, type of nursing program. In the fullest of these models (6.9), both the slopes and intercepts are allowed to vary, and type of nursing program predicts both the β_{0j}'s and the β_{1j}'s. In simpler forms of this model, such as in model 6.1, only the intercepts are allowed to vary and type of nursing program predicts the β_{0j}'s only. In model 6.2, only the slopes are allowed to vary while type of nursing program predicts the β_{0j}'s only. In model

6.3, both the slopes and intercepts are allowed to vary, and type of nursing program predicts the β_{0j}'s only. Models 6.4 through 6.6 follow the same pattern, except that type of nursing program predicts the β_{1j}'s only. Finally, in models 6.7 through 6.9, the intercepts and slopes follow the same pattern, and type of nursing program predicts both the β_{0j}'s and the β_{1j}'s. Second order PQL estimates for these models are found in Tables 13A through 13C.

Table 13A

Parameter Estimates for Model 6.1-6.3

	Model 6.1		Model 6.2		Model 6.3	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	2.058	.05339	2.081	.0528	2.081	.0528
τ_{00} (var(μ_{0j}))	0	0			0	0
γ_{10}	.04816	.00315	.04272	.00400	.04272	.00400
τ_{11} (var(μ_{1j}))			.00071	.00021	.00071	.00021
γ_{01}	-.00266	.09395	.1365	.1057	.1365	.1057
τ_{01} (cov(μ_{0j}/μ_{1j}))					0	0
Likelihood	3614.27		3190.52		3190.52	

Note. The fullest form of model 6 is specified as: $\text{Logit(NCLEX}_{ij}) = \gamma_{00} + \gamma_{01}\text{BACC}_j + \gamma_{10}\text{DRTC}_{ij} + \gamma_{11}\text{BACC}_j\text{DRTC}_{ij} + \mu_{0j} + \mu_{1j}\text{DRTC}_{ij}$

Table 13B

Parameter Estimates for Model 6.4-6.6

	Model 6.4		**Model 6.5**		**Model 6.6**	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	2.082	.0473	2.116	.0471	2.116	.0471
τ_{00} **(var(μ_{0j}))**	0	0			0	0
γ_{10}	.03887	.00370	.03575	.00439	.03575	.0044
τ_{11} **(var(μ_{1j}))**			.0005	.00018	.00049	.00018
γ_{11}	.0239	.00554	.02223	.00747	.02223	.00748
τ_{01} **(cov(μ_{0j}/μ_{1j}))**					0	0
Likelihood	3480.07		3199.69		3199.69	

Note. The fullest form of model 6 is specified as: Logit(NCLEX$_{ij}$) = $\gamma_{00} + \gamma_{01}$BACC$_j$ + γ_{10}DRTC$_{ij}$ + γ_{11}BACC$_j$DRTC$_{ij}$ + μ_{0j} + μ_{1j}DRTC$_{ij}$

Table 13C

Parameter Estimates for Model 6.7-6.9

	Model 6.7		**Model 6.8**		**Model 6.9**	
Parameter	est.	S.E.	est.	S.E.	est.	S.E.
γ_{00}	2.009	.05241	2.044	.05224	2.044	.05224
τ_{00} **(var(μ_{0j}))**	0	0			0	0
γ_{10}	.03741	.00369	.03504	.00426	.03504	.00426
τ_{11} **(var(μ_{1j}))**			.00041	.00017	.00041	.00017
γ_{01}	.3844	.1288	.3435	.1216	.3435	.1216
γ_{11}	.03761	.00738	.03419	.00858	.03419	.00858
τ_{01} **(cov(μ_{0j}/μ_{1j}))**					0	0
Likelihood	3553.67		3283.26		3283.26	

Note. The fullest form of model 6 is specified as: Logit(NCLEX$_{ij}$) = γ_{00} + γ_{01}BACC$_j$ + γ_{10}DRTC$_{ij}$ + γ_{11}BACC$_j$DRTC$_{ij}$ + μ_{0j} + μ_{1j}DRTC$_{ij}$

Models 7.1 - 7.9

The seventh family of models is similar to the sixth, except that the single level-one predictor included is centered PreAd score rather than centered DRT scores along with the level-two predictor, type of nursing program. In the fullest of these models (7.9), both the slopes and intercepts are allowed to vary, and type of nursing program predicts both the β_{0j}'s and the β_{1j}'s. In simpler forms of this model, such as in model 7.1, only the intercepts are allowed to vary and type of nursing program predicts the β_{0j}'s only. In model 7.2, only the slopes are allowed to vary while type of nursing program predicts the β_{0j}'s only. In model 7.3, both the slopes and intercepts are allowed to vary, and type of nursing program predicts the β_{0j}'s only. Models 7.4 through 7.6 follow the same pattern, except that type of nursing program predicts the β_{1j}'s only. Finally, in models 7.7 through 7.9, the intercepts and slopes follow the same pattern, and type of nursing program predicts both the β_{0j}'s and the β_{1j}'s. Second order PQL estimates for these models are found in Tables 14A through 14C.

Table 14A

Parameter Estimates for Model 7.1-7.3

	Model 7.1		Model 7.2		Model 7.3	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	1.983	.0544	1.982	.04748	1.966	.05162
τ_{00} (var(μ_{0j}))	.06463	.03267			.0392	.02966
γ_{20}	.01854	.00486	.01839	.00525	.01304	.00495
τ_{22} (var(μ_{2j}))			.00048	.00040	.00007	.00037
γ_{01}	-.2821	.09994	-.3051	.08455	-.2358	.0984
τ_{02} (cov(μ_{0j} /μ_{2j}))					-.0073	.00234
Likelihood	3682.57		3750.17		3769.55	

Note. The fullest form of model 7 is specified as: Logit(NCLEX$_{ij}$) = γ_{00} + γ_{01}BACC$_j$ + γ_{20}PreAdC$_{ij}$ + γ_{21}BACC$_j$PreAdC$_{ij}$ + μ_{0j} + μ_{2j}PreAdC$_{ij}$

Table 14B

Parameter Estimates for Model 7.4-7.6

	Model 7.4		Model 7.5		Model 7.6	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	1.915	.04764	1.899	.03987	1.908	.04547
τ_{00} (var(μ_{0j}))	.08244	.03502			.05951	.03262
γ_{20}	.01284	.00570	.0127	.00601	.01028	.00573
τ_{22} (var(μ_{2j}))			.0004	.0004	.00004	.00036
γ_{21}	.02172	.01064	.02439	.01146	.01152	.01108
τ_{02} (cov(μ_{0j}/μ_{2j}))					-.00716	.00244
Likelihood	3654.23		3784.31		3726.6	

Note. The fullest form of model 7 is specified as: $\text{Logit}(\text{NCLEX}_{ij}) = \gamma_{00} + \gamma_{01}\text{BACC}_j + \gamma_{20}\text{PreAdC}_{ij} + \gamma_{21}\text{BACC}_j\text{PreAdC}_{ij} + \mu_{0j} + \mu_{2j}\text{PreAdC}_{ij}$

Table 14C

Parameter Estimates for Model 7.7-7.9

	Model 7.7		Model 7.8		Model 7.9	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00}	1.976	.05435	1.975	.04742	1.969	.05169
τ_{00} (var(μ_{0j}))	.06472	.03266			.04004	.02982
γ_{20}	.01365	.005808	.014	.006168	.009236	.00583
τ_{22} (var(μ_{2j}))			.00040	.0004	.00001	.00035
γ_{01}	-.2497	.1027	-.2801	.08726	-.2801	.08726
γ_{21}	.01631	.01065	.01512	.01146	.01512	.01146
τ_{02} (cov(μ_{0j}/μ_{2j}))					-.00701	.00231
Likelihood	3669.09		3753.76		3762.45	

Note. The fullest form of model 7 is specified as: Logit(NCLEX$_{ij}$) = γ_{00} + γ_{01}BACC$_j$ + γ_{20}PreAdC$_{ij}$ + γ_{21}BACC$_j$PreAdC$_{ij}$ + μ_{0j} + μ_{2j}PreAdC$_{ij}$

C. Choice of the Final Model

In order to determine which model best fit the data, the models were compared using a variety of methods. The significance of each of the fixed parameters was easily determined by dividing the parameter estimate by its standard error, but as was discussed previously, the distribution of the ratio of the random parameters to their standard errors may depart from normal, so it has been suggested that the likelihood ratio statistic be used to compare models instead (Woodhouse, Rasbash, Goldstein, & Yang, 1995). However, examination of the tables in the previous section reveal that the likelihood values do not follow an expected pattern. For example, compare the likelihoods for models 3.1 and 3.3. In model 3.1, PreAdC is the only first level predictor, there are no second level predictors, the slopes are allowed to vary, and the intercepts are constrained. The likelihood for this model is approximately 3663. In model 3.3, both slopes and intercepts are allowed to vary, and with less restriction on the model one would expect that the likelihood would be as small as or smaller than the likelihood in model 3.1. However, as it turns out, the likelihood for this model is approximately 3744. This pattern of inconsistent likelihood values continues throughout the models. Goldstein and Rasbash (1995) have suggested that the likelihoods should not be used to make comparisons between discrete response models because the computation of the likelihood relies on the linearization approximation involved in the model estimation, and therefore comparisons may be unreliable. Consequently, it was decided that, in order to determine which model best fit the data, the fullest model possible would be examined, and non-significant predictors would be dropped. The fullest model possible is model 8.54, which can be found in Table 15, along with the second order PQL estimates for this model.

Table 15

Parameter Estimates for Model 8.54

	Model 8.54	
Parameter	**est.**	**S.E.**
γ_{00}	2.047	.05257
τ_{00} (var(μ_{0j}))	0	0
γ_{10}	.03467	.00431
τ_{11} (var(μ_{1j}))	.00041	.00017
γ_{20}	.00422 (n.s.)	.00621
τ_{22} (var(μ_{2j}))	0	0
γ_{01}	.3546	.1271
γ_{11}	.03339	.00864
γ_{21}	.00978 (n.s.)	.01202
τ_{01}(cov(μ_{0j} / μ_{1j}))	0	0
τ_{02}(cov(μ_{0j} / μ_{2j}))	0	0
τ_{12}(cov(μ_{1j} / μ_{2j}))	0	0
Likelihood	3294.17	

Note. The fullest form of model 8 is specified as: Logit(NCLEX$_{ij}$) = γ_{00} + γ_{01}BACC$_j$ + γ_{10}DRTC$_{ij}$ + γ_{20}PreAdC$_{ij}$ + γ_{11}BACC$_j$DRTC$_{ij}$ + γ_{21}BACC$_j$PreAdC$_{ij}$ + μ_{0j} + μ_{1j}DRTC$_{ij}$ + μ_{2j}PreAdC$_{ij}$

Examination of model 8.54 reveals that it is the fullest model since it contains all possible terms. Of the fixed terms, it was determined that PreAdC and BACCPreAdC could be dropped, since the parameter for the term divided by its standard error did not exceed 2. All of the random terms except τ_{11} were dropped because they were equal to zero. After dropping these terms, model 8.54 becomes equivalent to model 6.8 in which the slopes and intercepts were allowed to vary, and program type is included as a second level predictor. At the first level, the model can be specified as follows:

$$\text{Logit}(\text{NCLEX}_{ij}) = \beta_{0j} + \beta_{1j}\text{DRTC}_{ij}$$

A model that incorporates the school-level variable can be developed:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{BACC}_j$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\text{BACC}_j + \mu_{1j}$$

The parameter estimates obtained for this model are displayed in Table 16.

Table 16

Parameter Estimates for the Final Model

Parameter	est.	S.E.
γ_{00}	2.051	.05059
γ_{10}	.03462	.00428
γ_{01}	.3364	.1216
γ_{11}	.03371	.0086
τ_{11} (var(μ_{1j}))	.0005	.00018

Note. The final model is specified as: $Logit(NCLEX_{ij}) = \gamma_{00} + \gamma_{01}BACC_j + \gamma_{10}DRTC_{ij} + \gamma_{11}BACC_jDRTC_{ij} + \mu_{1j}DRTC_{ij}$

Substituting the parameter estimates into the equations for the more general model produces the following:

$$\beta_{0j} = 2.051 + .3364BACC_j$$

$$\beta_{1j} = .03462 + .03371BACC_j$$

Substituting the school-level equations into the student-level equation, the combined model becomes:

$$Logit(NCLEX_{ij}) = 2.051 + .3364BACC_j + (.03462 + .03371BACC_j)DRTC_{ij}$$

Multiplying out the terms, the combined model becomes:

$$Logit(NCLEX_{ij}) = 2.051 + .3364\ BACC_j + .03462\ DRTC_{ij} + .03371\ BACC_jDRTC_{ij}$$

In evaluating this model, we must first test the assumption that the residual variance is the same for baccalaureate programs as for non-baccalaureate programs (Braun, 1989). Contrast codes were developed to test this assumption, and the results indicated that there was no significant difference in the residual variances between the two types of programs (χ^2 (1) = 1.95, p=.163).

Another way to determine whether the model is appropriate for the data is to examine the residuals (Paterson, 1995). Standardized second-level residuals were calculated for the slope coefficients for the final model and plotted against values that would be expected from a standard Normal distribution. This plot is illustrated in Figure 1. If the residuals are close to being normally distributed, which would be the case if the model fit the data well, the plot should result in a relatively straight line. As can be seen from examining Figure 1, the residual plot for this model indicates that the data fit the model relatively well.

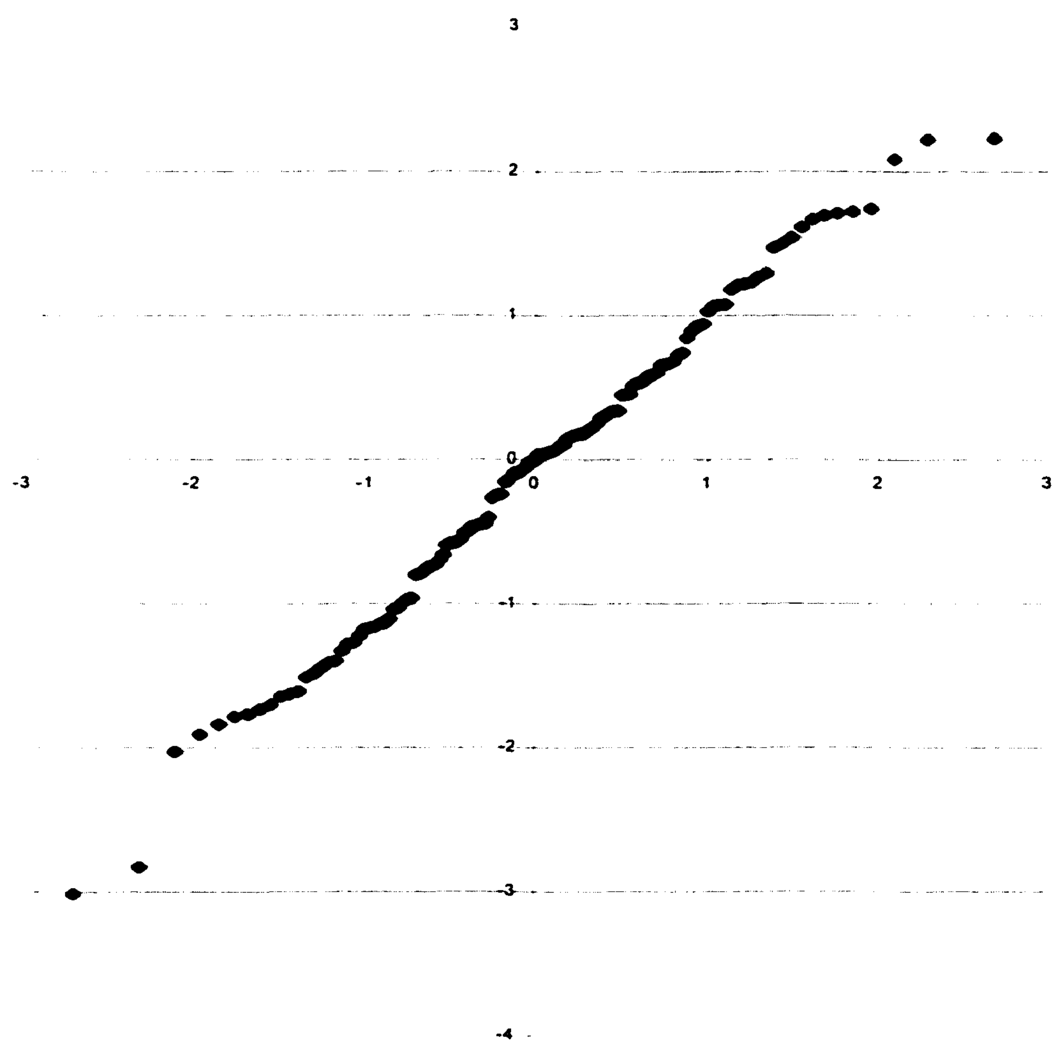

Figure 1. Residual plots for the final model.

A key question in this analysis was: does the predictive validity of the DRT vary as a function of type of nursing program? An examination of Table 16 reveals that the parameters that are associated with type of nursing program, γ_{01} and γ_{11}, are significant predictors of slope variability. Therefore, the predictive validity of the DRT must be examined separately by type of nursing program.

D. Differential Prediction

If we substitute the appropriate code for type of program, separate logistic regression equations can be developed for each of the two types of nursing programs. For example, if a "1" is substituted for BACC to reflect students in a baccalaureate program, the equation becomes:

$$\text{Logit}(\text{NCLEX}_{ij}) = 2.3874 + .06833\text{DRTC}_{ij}$$

If a "0" is substituted for BACC to reflect students in a non-baccalaureate program, the BACC portions of the equation drop out and the equation for non-baccalaureates simply becomes:

$$\text{Logit}(\text{NCLEX}_{ij}) = 2.051 + .03462\text{DRTC}_{ij}$$

The difference in these equations indicates that there is a different relationship between the DRT and the NCLEX-RN in each of the two populations, baccalaureate and non-baccalaureate. The larger slope for the baccalaureates indicate that the DRT is a stronger predictor of NCLEX-RN performance for them than for non-baccalaureates. We can examine the probability of passing the NCLEX-RN for students who obtain an average DRT score in each of the populations by substituting a "0" for DRTC score, since DRT was centered about the mean prior to the analysis. Doing so produces the following:

For baccalaureates:

Logit(NCLEX-RN) = 2.3874

The probability of passing the NCLEX-RN for baccalaureates who score at the mean on DRT is 91.6%. For non-baccalaureates, the equation becomes:

Logit(NCLEX-RN) = 2.051

So the probability of passing the NCLEX-RN for non-baccalaureates who score at the mean on DRT is only 88.6%. The slope differences provide evidence that the DRT is a stronger predictor of NCLEX-RN performance for the baccalaureates than for non-baccalaureates, but we have not yet addressed the question of the magnitude of that relationship for either program.

In establishing the predictive validity of an examination with a continuous criterion, the squared regression coefficient is a useful measure of the estimate of the validity, since it represents the proportion of variance in the criterion that is accounted for by the predictor. Unfortunately, there is no analogous measure in logistic regression analysis. However, we can examine the probability of passing the NCLEX-RN at different percentile ranks of DRT score for each of the program types. If the DRT is a significant predictor of NCLEX-RN performance, it should be able to discriminate between those who pass and those who fail the

NCLEX-RN, and so the probability of passing the NCLEX-RN should vary at different percentiles for each of the programs.

The 16th and 84th percentile scores were chosen for this portion of the analysis. These percentile scores should represent scores approximately one standard deviation above and below the mean if the scores are normally distributed. The 16th percentile score of the centered DRT scores is -14, while the 84th percentile score is 13. Substituting the percentile scores for DRTC in the equations for each of the types of nursing programs, and calculating the associated probability of passing the NCLEX-RN, produces the probabilities that are shown in Table 17.

Table 17

Probability of Passing the NCLEX-RN at the 16th and 84th Percentiles

	Baccalaureates	**Non-Baccalaureates**
16th Percentile	80.6%	82.6%
84th Percentile	96.3%	92.4%

Clearly, for students in both programs, students with lower DRT scores have a lower probability of passing the NCLEX-RN than those with higher DRT scores. However, the difference between the probability of passing the NCLEX-RN at the 16th and 84th percentile is smaller for the non-baccalaureates than for the baccalaureates, indicating that the DRT is a more valid predictor of NCLEX-RN performance for baccalaureate students.

E. Comparison of the HLM Analysis to Traditional Logistic Regression Analysis

Appendix A contains a listing of the traditional logistic regression coefficients for each school compared to the EB estimates that were obtained from the hierarchical analysis. Figure 2 demonstrates the relationship between the classical logistic regression coefficients for centered DRT scores, and the corresponding EB, or "shrinkage" estimators.

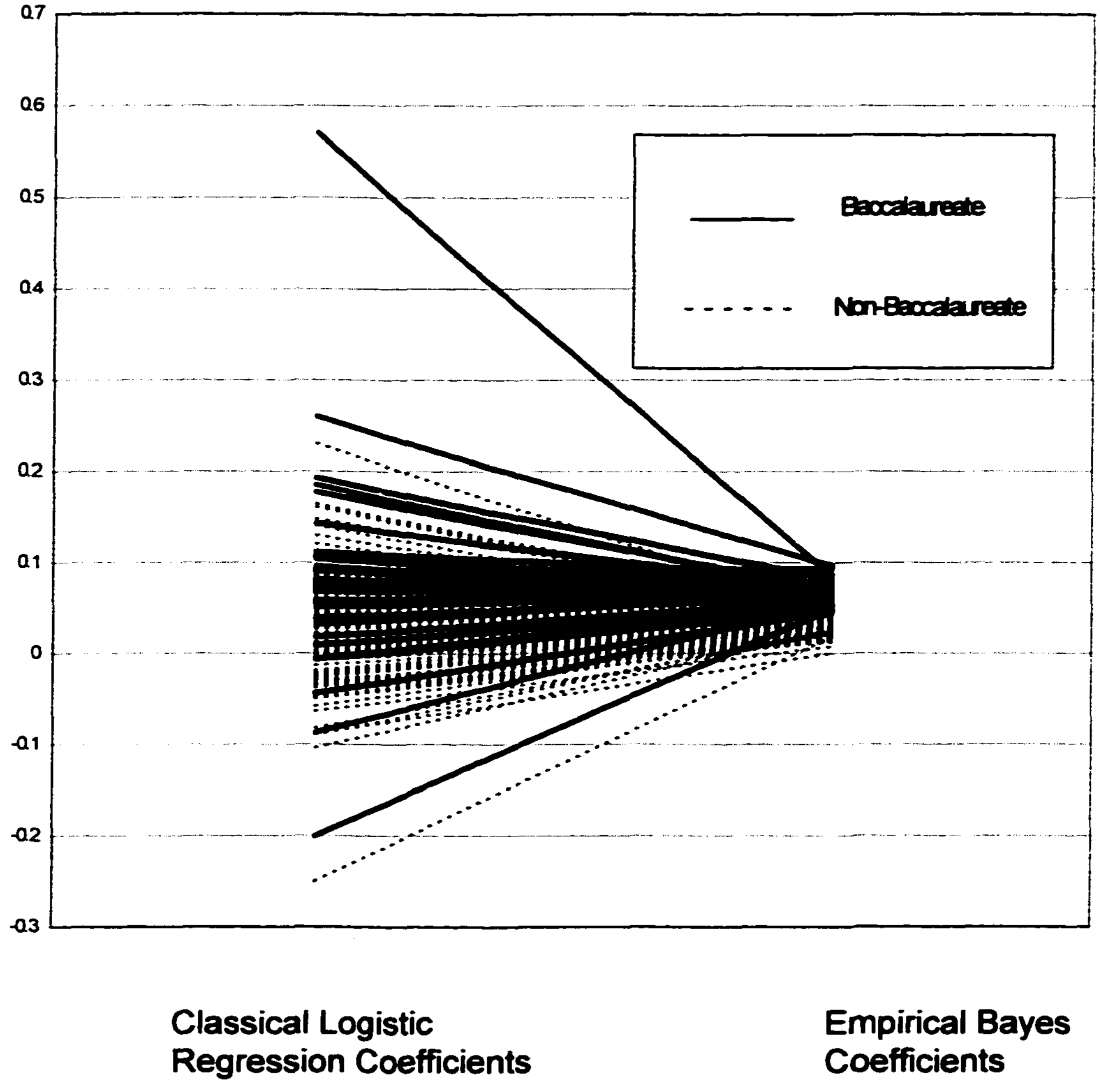

Figure 2. Shrinkage plot for the school-level slope coefficients related to DRTC scores.

In addition to comparing the coefficients, it is also interesting to examine the results of a traditional logistic regression analysis performed using the entire sample, simply including a dummy code for the type of nursing program and an interaction term to represent the BACC*DRTC term in order to simulate the final hierarchical model. Such a model would be specified as follows:

$$\text{Logit}(\text{NCLEX}_{ij}) = \beta_0 + \beta_1\text{DRTC}_{ij} + \beta_2\text{BACC}_j + \beta_3\text{BACC}_j\text{DRTC}_{ij}$$

Where:

$\text{Logit}(\text{NCLEX}_{ij})$	=	log-odds of passing NCLEX for student i in school j
β_0	=	intercept term, or the expected value of the log-odds of passing NCLEX for an non-baccalaureate individual for whom the value of DRTC is zero
β_1	=	slope for DRTC score, or the expected change in the log-odds of passing NCLEX for every unit change in DRTC
DRTC_{ij}	=	DRTC score for student i in school j
β_2	=	the expected difference in β_0 between baccalaureate students and non-baccalaureate students
BACC_j	=	an indicator variable to code program type; 1 if program type is baccalaureate, zero otherwise
β_3	=	the expected difference in DRTC slope between baccalaureate and non-baccalaureate students

$BACC_j{*}DRTC_{ij}$ = an term representing the interaction between DRTC score and baccalaureate/non-baccalaureate

The results of this analysis, with comparison to the hierarchical logistic regression analysis results, can be found in Table 18.

Table 18

Comparison of Classical Logistic Regression Results with Hierarchical Logistic Regression Results

	Hierarchical Logistic Regression Model		Classical Logistic Regression Model	
Parameter	**est.**	**S.E.**	**est.**	**S.E.**
γ_{00} or β_0	2.051	.0506	2.009	.0495
γ_{10} or β_1 (DRTC)	.0346	.0043	.0374	.0035
γ_{01} or β_2 (BACC)	.3364	.1216	.3841	.1212
γ_{11} or β_3 (BACC*DRTC)	.0337	.0086	.0376	.0070
τ_{11} (var(μ_{1j}))	.0004955	.0001764		

Using the same procedure as was used with the final hierarchical logistic regression model in terms of examining the probability of passing the NCLEX-RN at different percentile ranks of DRT scores for each of the program types, we obtain the following results from the classical logistic regression model, and compare them to previous results, in Table 19.

Table 19

Probability of Passing the NCLEX-RN at the 16th and 84th Percentiles for both the Classical and Hierarchical Models

	Hierarchical Logistic Regression Model		**Classical Logistic Regression Model**	
	Bacc	Non-Bacc	Bacc	Non-Bacc
16th Percentile	80.6%	82.6%	79.3%	81.5%
84th Percentile	96.3%	92.4%	96.7%	92.4%

CHAPTER V: DISCUSSION

A. Overview

The purpose of this study was twofold: to evaluate the predictive validity of the DRT using a hierarchical logistic regression model and to compare these results to a more traditional logistic regression approach. In evaluating both the hierarchical and the traditional models using the entire sample, it is clear that both sets of results indicate that there was significant school-level variability in the slopes that could be explained by type of nursing program. These results indicate that, for the national sample of nursing students, the predictive validity of the DRT should be reported separately for baccalaureate and non-baccalaureate programs. When examining the results of the hierarchical analysis, however, it also became clear that the EB coefficients obtained were superior to the coefficients obtained using the traditional approach in that none of the EB slope estimates were negative. Indeed, only the results of the hierarchical analysis support the generalizability of the validity findings by type of nursing program. The shrinkage estimators shown in Figure 2 indicate that the EB estimates by individual institution cluster strongly around type of institution, and so it can be argued that institutions need not report separate validity coefficients. These findings have implications for both nursing education practice and policy as well as for the application of multilevel techniques in predictive validity problems.

These implications are discussed more fully in the following sections. The discussion sections ends with a review of the study's limitations and suggestions for further research.

B. Implications for Nursing Education

As was discussed earlier, there has been quite a lot of contention in the nursing education community regarding the 1965 recommendation of the American Nursing Association that the baccalaureate degree be the minimum credential for entry into practice. The data presented in this study, obtained from a national sample of nursing licensure candidates, support earlier findings that associate degree and diploma graduates pass the NCLEX-RN at a higher rate than do baccalaureate graduates. In addition, the predictive validity of the DRT, which is a test designed to be as parallel as possible to the NCLEX-RN, is significantly higher for candidates enrolled in baccalaureate programs than for candidates enrolled in non-baccalaureate programs. This finding suggests that the ways in which the DRT and similar tests are used instructionally by different types of nursing programs may have an effect on the ability of such tests to predict future performance on the licensure exam. The DRT provides a great deal of diagnostic information to the student in order to help identify her areas of strength and weakness, and to help her prepare for the NCLEX-RN. Ironically, when such a diagnostic test is used appropriately, it would be expected that its predictive validity would be lower than if the diagnostic information were not used instructionally before the criterion measure was administered. This point will be discussed further in the next section on psychometric implications. However, to explore how this type of validity coefficient depression might be playing out in the field of nursing education, consider the following hypothesis: it may be that faculty in non-baccalaureate nursing programs actively use the results of the DRT in preparing students for the NCLEX-RN than do faculty in baccalaureate programs. If true, we would expect that faculty in non-baccalaureate programs would strive to address identified areas of student weakness with specific instruction, so that subsequent NCLEX-RN performance for a particular student would be higher than that student's DRT score would have predicted. This, in turn, would depress the predictive validity of the DRT.

Circumstantial evidence regarding the differential use that is made of the DRT suggests that this may be what is happening in the nursing education community. The DRT can be sold either to individual students or to an institution for classroom administration. Sales of the DRT indicate that non-baccalaureate programs are more likely to purchase the test institutionally and administer the test during regular classroom time, while baccalaureate students are more likely to purchase the test individually (NLN, personal communication). This in turn suggests that the non-baccalaureate programs may also use the diagnostic test results during regular classroom time to provide guided instruction in the identified areas of weakness. Baccalaureate students, on the other hand, may very well be left to their own devices in using the results of the DRT to prepare for the NCLEX-RN. One possible recommendation for the baccalaureate programs that emerges from this analysis is that they use tests like the DRT more proactively as they prepare students to sit for the licensure exam.

However, proponents of baccalaureate education may very well carp at such a recommendation. One of the arguments that has been made in support of baccalaureate nursing education is that higher-level skills such as decision making and critical thinking, which are crucial to sound professional practice, can only be effectively taught at the this level (Riffle, et al., 1985). Indeed, these advocates have been highly critical of the intensive training and testing of basic skills that have been a hallmark of associate degree programs (Riffle, et al., 1985). Given the clear distinctions between baccalaureate and non-baccalaureate programs, both in terms of their philosophical underpinnings and in the empirical evidence presented here and in other studies, a more effective model for nursing credentialling may be a multiple examination type of licensure system that allows the two groups to demonstrate their different proficiencies. A model that incorporates some of the features of the Praxis system for education credentialling may be a useful model for nurses as well (Educational Testing Service, 1998). The Praxis series, which is used in various combinations in 34 states, consists of three distinct sets of assessments. Praxis I is designed to measure pre-professional skill levels in reading, writing, and mathematics, much like the NLN's Pre-Admission test, and is used for admission into teacher preparation programs. Praxis II consists of subject matter assessments that are used by many states for initial licensure. Praxis III assessments evaluate the actual classroom performance of beginning teachers. These assessments are administered during the first year of teaching, and are often used to make more permanent licensure decisions. A similar system could be developed for the nursing profession, in which an initial licensure examination, such as the NCLEX-RN, is supplemented with a performance assessment administered after a period of adjustment. Baccalaureate nursing graduates might very well out-perform their non-baccalaureate counterparts on the performance assessment as they bring their higher-level skills to bear upon complex, real-world situations.

Alternatively, some states may want to pursue distinguishing the baccalaureate from the non-baccalaureate practitioner by credentialling the baccalaureate nursing graduate at a higher level and thereby creating a two-tiered licensure system, similar to the existing distinction that is made between L.P.N.'s (licensed practical nurses) and R.N.'s. In order to validate such a system, however, it would have to be demonstrated that baccalaureate graduates could be hired to perform duties that are distinctly different from the non-baccalaureates upon graduation. In addition, it would be a difficult alternative to pursue politically in light of the perceived benefits of allowing relatively easy access to a respected and well-remunerated profession.

C. Implications for Psychometric Applications of HLM

The results of this study indicate that it is indeed useful to apply hierarchical logistic regression modeling techniques to predictive validity studies with a dichotomous criterion, particularly when there is interest in providing accurate coefficient estimates for each of the level-two units. Although similar results were obtained using a traditional approach for the overall sample, the school-level EB coefficients estimates obtained in the hierarchical approach appear to be more accurate and "sensible", in that none of them were negative. In addition, the results obtained from the pilot indicated that there was a great deal of variability in the school-level coefficients, but there was no way of determining how much of the variability was due to random sampling error. Using a multilevel model allowed the direct modeling of these random effects, resulting in more accurate school-level estimates. As was pointed out previously, it is oftentimes crucial to be able to provide this school-level data so that institutions can decide on the utility of a particular test for use with their own population of students. It could also be argued that, even though the overall results of the hierarchical analysis were not that different from the more traditional analysis, it was necessary to conduct the hierarchical study in order to discover that. There has not been enough work done in the field of multilevel modeling as yet that would help determine a priori which studies involving nested data might not benefit from a hierarchical treatment.

Even as an improvement over a traditional approach, the question of whether an EB treatment is "good enough" may still arise. Some researchers may argue for a fully Bayesian analysis rather than an EB approach. However, researchers who have compared EB approaches to fully Bayesian ones have found that the inferences made from the results of EB analyses are not substantively different from those made from a Bayes approach (Braun, 1989; Rubin 1981).

This study has also suggested several ways in which an estimate of the predictive validity can be illustrated when the criterion is dichotomous. In these types of studies, there is nothing immediately analogous to the multiple correlation coefficient that is obtained with a continuous criterion. The results therefore were examined in several ways. One method involved comparing and contrasting the probability of passing the NCLEX-RN at different percentile ranks of DRT score. There were clear differences between these probabilities for both types of programs, indicating that the DRT was indeed predictive of NCLEX-RN performance, although, as expected, the differences were smaller for non-baccalaureate programs.

Another area where this work is helpful in terms of expanding available analytic techniques is the area of validity generalization. Traditionally studied in the area of industrial/organizational psychology, validity generalization approaches attempt to identify, usually through meta-analytic techniques, workplace and employee characteristics that would allow employment screening or certification/licensure instruments to be validly used across similar institutions (Burke, 1984; Kane, 1990; Rafilson, 1991). While individual institutions may prefer to obtain individualized validity coefficients, providing them can be problematic. When an institution has a small and/or highly selective population, the validity coefficient may not be very accurate. An institution may also want to use a new instrument without necessarily first collecting pilot data. Using a multilevel approach, however, provides a framework for calculating validity coefficients based on important institution- and student-level variables, and allows the generalization of validity coefficients across similar types of institutions. The results obtained in this study, for example, indicate that validity coefficients could be reported by type of nursing program - rather than by individual school - without loss of information.

A troubling aspect of reporting these predictive validity coefficients for the DRT, which is a diagnostic assessment, is the coefficient depression problem that was mentioned in the previous section. The appropriate use of a diagnostic test may result in a depressed predictive validity coefficient if instruction is tailored to the diagnostic information before the criterion measure is administered. However, validity models for these types of assessments will need to be reconceptualized as diagnostic assessments become technically more sophisticated and more readily available. Several years ago, Tittle (1989, 1994) described the need for better models of the links between teaching practice and assessment theory (Tittle, 1994), and pointed out that "development of educational assessments must take place within an understanding of how teachers and students can and do use such information. Some research evidence supports the need to expand validation inquiry to evaluate test use in context, in the situations and information systems in which the assessments are embedded." (Tittle, 1989, p. 13). The present study illustrates this need to understand the context in which the assessment results are used in order to make sense of the predictive validity results. However, it can be argued that predictive validity models, in their current conceptualization, may not be an optimal way of assessing the technical qualities of a diagnostic instrument. It may be possible to specify what the predictive validity of a diagnostic instrument would be if used as intended by collecting more data from the end-user's perspective.

D. Suggestions for Further Research

Clearly, this study has limitations that must be taken into account when evaluating its results. Although the sample obtained was a national sample, and there did not seem to be any significant differences between the candidates for whom NCLEX-RN results were obtained and those for whom results were not obtained, there may be other differences between the groups that may have affected the results. In addition, a limited number of variables were available, so that particular types of issues that are currently of interest in the study of predictive validity, such as subgroup differences, could not be explored. The scarcity of school-level explanatory variables, together with the observations made in the last section that addressed the need for more contextual information in evaluating predictive validity studies, speaks to the need for detailed case studies of various types of institutions with varying pass rates on the NLCEX-RN. However, as this study represents an initial attempt to explore the applicability of hierarchical data analytic techniques to a particular kind of predictive validity problem, the results support this applicability and suggest further avenues of research. Certainly, the techniques presented in this study could be easily expanded for use with high-stakes, large-scale assessments, such as the SAT or GRE. There is evidence that post-secondary admissions officers take school-level differences, such as quality of high school curriculum, into account in a non-systematic way when comparing candidates for admission (Willingham, 1990). An application of HLM techniques, taking these school-level variables into account in a more systematic way, would help the admissions profession enormously. In addition, validity generalization studies could be conducted using EB estimates to overcome the problem that many post-secondary institutions have in calculating least squares estimates using small, restricted samples.

In addition, as was pointed out previously, many licensure and certification examination results are reported as 'pass/fail' rather than on a continuous scale.

Many organizations that sponsor such examinations are extremely interested in better modeling the factors that predict success (Craig Mills, personal communication), and the results presented here provide a framework for conducting similar studies.

Appendix A

Table Comparing Traditional Logistic Regression Estimates and Empirical Bayes Estimates by School

School	Program Type	Constant EB	Constant Traditional Logistic Regression	DRT Coefficient EB	DRT Coefficient Logistic Regression
1	Non-Bacc	2.051	3.0247	0.045479	.1211
2	Non-Bacc	2.051	2.0102	0.023678	-.0023
3	Non-Bacc	2.051	2.0799	0.024193	-.0214
4	Non-Bacc	2.051	2.2984	0.039772	.0812
5	Non-Bacc	2.051	5.6068	0.018246	-.2502
6	Bacc	2.3874	2.2105	0.065131	.0518
7	Non-Bacc	2.051	2.3081	0.035755	.0454
8	Non-Bacc	2.051	2.7453	0.032336	-.0069
9	Non-Bacc	2.051	1.6389	0.031102	.0131
10	Non-Bacc	2.051	2.1715	0.066445	.1113
11	Non-Bacc	2.051	1.514	0.024358	-.0063
12	Non-Bacc	2.051	2.1234	0.061354	.0823
13	Non-Bacc	2.051	2.69	0.040688	.0974
14	Non-Bacc	2.051	3.1875	0.04643	.1647
15	Non-Bacc	2.051	0.9559	0.036433	.0281
16	Bacc	2.3874	6.0562	0.096608	.2602
17	Bacc	2.3874	3.5519	0.063376	.0812
18	Non-Bacc	2.051	1.527	0.03173	.0426
19	Non-Bacc	2.051	2.0468	0.035415	.0372
20	Non-Bacc	2.051	3.7561	0.025067	-.1037
21	Non-Bacc	2.051	2.7148	0.040618	.0731
22	Non-Bacc	2.051	1.4572	0.028826	-.0186
23	Bacc	2.3874	3.4713	0.056657	-.0868
24	Non-Bacc	2.051	2.4354	0.022829	-.0820
25	Non-Bacc	2.051	1.6425	0.03079	.0104
26	Non-Bacc	2.051	2.535	0.031079	-.0041
27	Non-Bacc	2.051	2.4567	0.05201	.0845
28	Non-Bacc	2.051	1.8342	0.039231	.0427
29	Non-Bacc	2.051	2.6994	0.027841	.0316
30	Non-Bacc	2.051	1.5975	0.034244	.0387
31	Non-Bacc	2.051	1.4471	0.038312	.0272

School	Program Type	Constant EB	Constant Traditional Logistic Regression	DRT Coefficient EB	DRT Coefficient Logistic Regression
32	Non-Bacc	2.051	2.1359	0.0357	.0427
33	Non-Bacc	2.051	3.0335	0.050521	.0959
34	Bacc	2.3874	4.6758	0.081709	.1933
35	Non-Bacc	2.051	2.8236	0.042805	.1111
36	Non-Bacc	2.051	2.6586	0.024381	-.0479
37	Non-Bacc	2.051	1.9114	0.027226	-.0008
38	Non-Bacc	2.051	2.2026	0.034702	.0392
39	Non-Bacc	2.051	3.6201	0.020579	-.0579
40	Bacc	2.3874	2.0279	0.063832	.0384
41	Bacc	2.3874	2.9092	0.071054	.0967
42	Non-Bacc	2.051	2.2829	0.051831	.0908
43	Non-Bacc	2.051	2.3256	0.017217	-.0422
44	Non-Bacc	2.051	2.2106	0.019864	-.0375
45	Bacc	2.3874	2.2221	0.065964	.0578
46	Non-Bacc	2.051	1.7009	0.036766	.0344
47	Non-Bacc	2.051	2.3073	0.03633	.0532
48	Non-Bacc	2.051	1.9774	0.019142	-.0335
49	Non-Bacc	2.051	2.4154	0.025884	.0090
50	Non-Bacc	2.051	1.9714	0.03525	.0371
51	Bacc	2.3874	2.1036	0.075411	.0782
52	Bacc	2.3874	2.6829	0.06621	.0740
53	Non-Bacc	2.051	1.1343	0.038801	.0408
54	Non-Bacc	2.051	2.4924	0.000258	-.0881
55	Non-Bacc	2.051	2.7823	0.032318	.0228
56	Bacc	2.3874	2.1305	0.093213	.0913
57	Bacc	2.3874	2.864	0.077329	.1048
58	Non-Bacc	2.051	3.3254	0.047205	.1616
59	Non-Bacc	2.051	1.7367	0.05275	.1061
60	Non-Bacc	2.051	1.4578	0.033463	.0220
61	Bacc	2.3874	2.9535	0.069	.0917
62	Bacc	2.3874	4.5287	0.048609	-.2001
63	Non-Bacc	2.051	1.9361	0.034429	.0291
64	Non-Bacc	2.051	1.5129	0.035922	.0277
65	Bacc	2.3874	2.428	0.073634	.0785
66	Non-Bacc	2.051	2.8929	0.047398	.0929
67	Non-Bacc	2.051	2.2245	0.038747	.0537
68	Non-Bacc	2.051	2.1364	0.030867	.0034

School	Program Type	Constant EB	Constant Traditional Logistic Regression	DRT Coefficient EB	DRT Coefficient Logistic Regression
69	Bacc	2.3874	2.7416	0.052467	.0094
70	Non-Bacc	2.051	2.0201	0.035107	.0371
71	Non-Bacc	2.051	2.01	0.051702	.0856
72	Non-Bacc	2.051	2.3318	0.024179	-.0354
73	Non-Bacc	2.051	2.5546	0.035005	.0014
74	Non-Bacc	2.051	1.5725	0.020563	-.0341
75	Bacc	2.3874	1.0315	0.053841	-.0059
76	Bacc	2.3874	2.9528	0.067502	.0941
77	Bacc	2.3874	4.6733	0.070959	.1778
78	Bacc	2.3874	1.5495	0.066916	.0324
79	Non-Bacc	2.051	2.4447	0.027514	-.0069
80	Non-Bacc	2.051	2.4265	0.036511	.0475
81	Non-Bacc	2.051	1.7452	0.024792	-.0358
82	Bacc	2.3874	2.2271	0.062001	.0394
83	Bacc	2.3874	1.9244	0.067514	.0568
84	Bacc	2.3874	2.0819	0.045861	-.0437
85	Non-Bacc	2.051	2.5373	0.02207	-.0227
86	Non-Bacc	2.051	2.2901	0.04111	.0619
87	Non-Bacc	2.051	1.7201	0.028633	.0013
88	Non-Bacc	2.051	2.4611	0.041708	.0706
89	Non-Bacc	2.051	2.0073	0.020774	-.0249
90	Non-Bacc	2.051	2.573	0.044618	.0676
91	Non-Bacc	2.051	2.6265	0.043143	.0817
92	Non-Bacc	2.051	1.9086	0.040757	.0522
93	Non-Bacc	2.051	2.2069	0.017103	-.0312
94	Non-Bacc	2.051	2.4412	0.030807	-.0035
95	Bacc	2.3874	2.079	0.069052	.0579
96	Bacc	2.3874	5.3548	0.068078	.1856
97	Bacc	2.3874	3.7106	0.059694	.1090
98	Non-Bacc	2.051	2.8023	0.036531	.0793
99	Non-Bacc	2.051	1.9205	0.007968	-.0634
100	Non-Bacc	2.051	1.443	0.055469	.0660
101	Non-Bacc	2.051	5.1026	0.048192	.2308
102	Non-Bacc	2.051	1.5148	0.059682	.1092
103	Non-Bacc	2.051	3.2147	0.027838	.0177
104	Bacc	2.3874	9.8453	0.085686	.5702
105	Bacc	2.3874	3.7157	0.071973	.1435

School	Program Type	Constant EB	Constant Traditional Logistic Regression	DRT Coefficient EB	DRT Coefficient Logistic Regression
106	Non-Bacc	2.051	2.9736	0.020779	-.0483
107	Non-Bacc	2.051	2.0953	0.058452	.1483
108	Bacc	2.3874	2.7002	0.086533	.1116
109	Non-Bacc	2.051	2.1982	0.014477	-.0475
110	Non-Bacc	2.051	1.2528	0.041266	.0514
111	Bacc	2.3874	1.8501	0.08686	.0791
112	Bacc	2.3874	2.3974	0.06824	.0686
113	Non-Bacc	2.051	2.4103	0.048822	.0791
114	Non-Bacc	2.051	3.0113	0.014493	-.0494
115	Non-Bacc	2.051	3.7781	0.041811	.0801
116	Non-Bacc	2.051	2.4344	0.038294	.0464
117	Non-Bacc	2.051	2.1676	0.030658	.0292
118	Non-Bacc	2.051	2.3226	0.016037	-.0343
119	Non-Bacc	2.051	2.615	0.040642	.0792
120	Bacc	2.3874	2.4272	0.057657	.0183
121	Bacc	2.3874	1.8191	0.046283	-.0062
122	Non-Bacc	2.051	2.6316	0.029274	-.0263
123	Non-Bacc	2.051	3.5285	0.040287	.1413
124	Non-Bacc	2.051	1.7183	0.025483	-.0128
125	Non-Bacc	2.051	3.797	0.024837	-.0895
126	Non-Bacc	2.051	1.3842	0.042755	.1303
127	Non-Bacc	2.051	1.9909	0.057931	.0882
128	Non-Bacc	2.051	1.9592	0.056913	.0832
129	Non-Bacc	2.051	1.5664	0.046601	.0517
130	Non-Bacc	2.051	3.1504	0.053113	.1648
131	Non-Bacc	2.051	1.488	0.029454	.0184
132	Non-Bacc	2.051	1.8664	0.025179	-.0059
133	Non-Bacc	2.051	1.8839	0.027427	.0069
134	Non-Bacc	2.051	1.75	0.013032	-.0293
135	Non-Bacc	2.051	3.489	0.036799	-.0210

REFERENCES

Aiken, L.S., & West, C.H. (1991). Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage.

Akers, S.F. (1992). Predicting program completion and success on the National Council Licensure Examination for nursing students in Mississippi. Unpublished doctoral dissertation, Mississippi State University, Mississippi.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: Author.

Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.

Beitz, J.M. (1994). Outcome assessment: Choosing a psychometrically sound nursing achievement test. Nurse Educator, 19(4), 12-17.

Bergstrom, B., Gershon, R., & Lunz, M.E. (1994). Exploring examinee response time using hierarchical linear modeling. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Braun, H.I. (1989). Empirical Bayes methods: A tool for exploratory analysis. In R.D. Bock (Ed.), Multilevel Analysis of Educational Data (pp. 19-55). New York, NY: Academic Press.

Braun, H.I., Jones, D.H., Rubin, D.B., & Thayer, D.T. (1983). Empirical Bayes estimation of coefficients in the general linear model from data of deficient rank. Psychometrika, 48, 171-181.

Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101, 147-158.

Bryk, A.S., & Raudenbush, S.W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

Burke, M.J. (1984). Validity generalization: A review and critique of the correlation model. Personnel Psychology, 37(1), 93-115.

Burton, B. (1993). Some observations on the effect of centering on the results obtained from hierarchical linear modeling. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Campbell, A.R., & Dickson, C.J. (1996). Predicting student success: A 10-year review using integrative review and meta-analysis. Journal of Professional Nursing, 12(1), 47-59.

Campbell, M.I. (1988). The relationships of clinical experience student-faculty ratio, clinical experience instructional activities, curricular content, and educational environment to Ohio schools of nursing graduate achievement on the National Council Licensure Examination. Unpublished doctoral dissertation, Ohio University, Athens.

Chornick, N.L., & Yocom, C.J. (1995). NCLEX job analysis study: Questionnaire development. Journal of Nursing Education, 34(3), 101-105.

Cronbach, L.J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), Test Validity (pp. 3-17). Hillsdale, NJ: Erlbaum.

Drake, C.C., & Michael, W.B. (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(4), 675-683.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. Journal of Educational and Behavioral Statistics, 20, 115-148.

Draper, D., Gaver, D.P., Goel, P.K., Greenhouse, J.B., Hedges, L.V., Morris, C.N., Tucker, J.R., & Waterneaux, C. (1993). Combining information: Statistical issues and opportunities for research. Alexandria, VA: American Statistical Association.

Duffy, M.E. (1990). Needed: New models for professional nursing education. In N.L. Chaska (Ed.), The Nursing Profession (pp. 84-92). St. Louis, MO: C.V. Mosby.

Educational Testing Service (1998). The Praxis series: 1998-99 registration bulletin. Princeton, NJ: Author.

Efron, B., & Morris, C. (1975). Data analysis using Stein's estimator and its generalization. Journal of the American Statistical Association, 70, 311-319.

Foti, I., & DeYong, 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.

Fitzpatrick, J.J., & Modly, D.M. (1990). The first doctor of nursing (ND) program. In N.L. Chaska (Ed.), The Nursing Profession (pp. 93-99). St. Louis, MO: C.V. Mosby.

Goldstein, H. (1987). Multilevel models in educational and social research. New York, NY: Oxford University Press.

Goldstein, H., & McDonald, R.P. (1988). A general model for the analysis of multilevel data. Psychometrika, 53, 455-467.

Goldstein, H., & Rasbash, J. (1995). Improved approximations for multilevel models with binary responses. [on-line]. Available: http://www.medent.umontreal.ca/multilevel/faq-mln.html#discretelike.

Gross, Y.T., Takazawa, E.S., & Rose, C.L. (1987). Critical thinking and nursing education. Journal of Nursing Education, 26(8), 317-323.

Homser, D.W., & Lemeshow, S. (1989). Applied Logistic Regression. New York, NY: John Wiley & Sons.

House, D.J., & Johnson, J.J. (1993). Graduate Record Examination scores and academic background variables as predictors of graduate degree completion. Educational and Psychological Measurement, 53(2), 551-556.

Kane, M.T. (1990). Generalizing criterion-related validity evidence for certification requirements across situations and specialty areas. (ACT Research Report 90-3). Iowa City, IA: ACT, Inc.

Karns, M.E.H. (1988). Nursing programs and state licensure examination performance: A comparative study of baccalaureate, associate degree, and diploma programs. Unpublished doctoral dissertation, University of Pittsburgh.

Kreft, I.G.G., & de Leeuw, J. (1998). Introducing multilevel modeling. Newbury Park, CA: Sage.

Kreft, I.G.G., de Leeuw, J., & Aiken, L.S. (1995). The effect of different forms of centering in hierarchical linear models. Multivariate Behavioral Research, 30(1), 1-21.

Lee, V.E., & Bryk, A.S. (1989). A multilevel model of the social distribution of high school achievement. Sociology of Education, 62,172-192.

Lindley, D.V., & Smith, A.F.M. (1972) . Bayes estimates for the linear model (with discussion). Journal of the Royal Statistical Society, Series B, 34, 1-34.

Mason, W.M., Wong, G.M., & Entwistle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), Sociological Methodology (pp. 72-103). San Francisco, CA: Jossey-Bass.

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

Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-103). New York, NY: Macmillan.

Messick, S. (1995). Validity of psychological assessment. American Psychologist, 50(9), 741-749.

National League for Nursing (1984). Self-study manual, council of baccalaureate and higher degree programs, Pub. No. 15-1955. New York, NY: Author.

National League for Nursing (1988). Diagnostic Readiness Test for RN Licensure. New York, NY: Author.

National League for Nursing (1989). NLN Diagnostic Readiness Test for RN Licensure User's Guide. New York, NY: Author.

National League for Nursing Press (1995). Nursing Data Review 1995. New York, NY: Author.

Oermann, M.H. (1991). Professional nursing practice: A conceptual approach. Philadelphia, PA: Lippincott.

Paterson, L. (1995). Entry to university by school leavers. In G. Woodhouse (Ed.), A Guide to Mln for New Users (pp. 87-101). London: Institute of Education, University of London.

Perez, T.L. (1977). Investigation of academic moderator variables to predict success on State Boards of Nursing in a baccalaureate nursing program. Journal of Nursing Education, 16, 16-23.

Raju, N.S., Steinhaus, S.D., Edwards, J.E., & DeLessio, J. (1993). A logistic regression model for personnel selection. Applied Psychological Measurement, 15(2), 139-152.

Rafilson, F. (1991). The case for validity generalization. (ERIC Document Reproduction Service No. ED 338 699)

Rasbash, J., & Woodhouse, G. (1996). Mln Command Reference, v1.0a. London: Institute of Education, University of London.

Raudenbush, S.W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. Journal of Educational Psychology, 76, 85-97.

Raudenbush, S.W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. Journal of Educational Statistics, 18(4), 321-349.

Records, N.L., & Tomblin, B.J. (1994). Clinical decision making: Describing the decision rules of practicing speech-language pathologists. Journal of Speech and Hearing Research, 37(1), 144-156.

Reed, C., & Feldhusen, J. (1972). State Board examination prediction for associate degree nursing program graduates. Nursing Research, 21, 149-153.

Riffle, K., Lamberth, F., Moine, G., & Fielding, J. (1985). Entry into practice: The continuing debate. In J.C. McClosky & H.K. Grace (Eds.), Current Issues in Nursing (2nd ed., pp. 33-75). Boston, MA: Blackwell Science.

Rogosa, D., & Saner, H. (1995). Longitudinal data analysis examples with random coefficient models. Journal of Educational and Behavioral Statistics, 20, 149-170.

Rubin, D.B. (1980). Using empirical Bayes techniques in the law school validity studies. Journal of the American Statistical Association, 75, 801-816.

Rubin, D.B. (1981). Estimation in parallel randomized experiments. Journal of Educational Statistics, 6(4), 377-400.

Rubin, D.B. (1989). Some applications of multilevel models to educational data. In R.D. Bock (Ed.), Multilevel Analysis of Educational Data (pp. 1-17). New York, NY: Academic Press.

Sakalys, J.A., & Watson, J. (1986). Professional education: Post-Baccalaureate education for professional nursing. Journal of Professional Nursing, 2, 91-97.

Saxton, D.F., Pelikan, P.K., Nugent, P.M., & Needleman, S.R. (1989). Mosby's Assess Test: A Practice Exam for RN Licensure. St. Louis, MO: CV Mosby.

Schmidt, A.E. (1997). An approximation of a hierarchical logistic regression model used to establish the predictive validity of an examination with a dichotomous criterion measure. Paper presented at the annual meeting of the National Council of Measurement in Education, Chicago, IL.

Schumacker, R.E., & Bembry, K. (1995). Centering effects in HLM level-I predictor variables. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Spahr, A. E. (1995). Predicting graduation status of nursing students using entering GPA and grades in algebra, biology, and chemistry. (Report No. 143). Cicero, IL: Morton College. (ERIC Document Reproduction Service No. ED 386 224)

Tittle, C.K. (1989). Validity: Whose construction is it in the teaching and learning context? Educational Measurement: Issues and Practices, 8, 5-13, 34.

Tittle, C.K. (1994). Toward an educational psychology of assessments for teaching and learning: Theories, contexts, and validation arguments. Educational Psychologist, 29(3), 149-162.

Vancouver, J.B., Millsap, R.E., & Peters, P.A. (1994). Multilevel analysis of organizational goal congruence. Journal of Applied Psychology, 79(5), 666-679.

Waterhouse, J.K., Carroll, M.C., & Beekman, P.B. (1993). National Council Licensure Examination success: Accurate prediction of student performance on the post-1988 examination. Journal of Professional Nursing, 9(5), 278-283.

Williamson, J.A. (1983). Crisis in academic nursing. In N.L. Chaska (Ed.), The Nursing Profession: A time to speak (pp. 63-69). New York, NY: McGraw Hill.

Willingham, W.W. (1990). Interpreting predictive validity. In W.W. Willingham, C. Lewis, R. Morgan, & L. Ramist (Eds.) Predicting college grades: An analysis of institutional trends over two decades (pp. 3-21). Princeton, NJ: Educational Testing Service.

Woodhouse, G., Rasbash, J., Goldstein, H., & Yang, M. (1995). Introduction to multilevel modeling. In G. Woodhouse (Ed.), A Guide to Mln for New Users (pp. 9-57). London: Institute of Education, University of London.

Wong, G.Y., & Mason, W.M. (1985). The hierarchical logistic regression model for multilevel analysis. Journal of the American Statistical Association, 80, 513-524.

Zwick, R. (1993). The validity of the GMAT for the prediction of grades in doctoral study in business and management: An empirical Bayes approach. Journal of Educational Statistics, 18, 91-107.