Ron and Kathy Assaf College of Nursing Student Theses, Dissertations and Capstones

Ron and Kathy Assaf College of Nursing

1-1-2019

Use of Evidence-Based Test Development in Pre-Licensure Nursing programs: A Descriptive Study of Faculty Beliefs, Attitudes and Values

Richild Berrick
Nova Southeastern University

This document is a product of extensive research conducted at the Nova Southeastern University College of Nursing. For more information on research and degree programs at the NSU College of Nursing, please click here.

Follow this and additional works at: https://nsuworks.nova.edu/hpd_con_stuetd

Part of the Nursing Commons

The Use of Evidence-Based Test Development in Pre-Licensure Nursing Programs:
A Descriptive Study of Faculty Beliefs, Attitudes and Values

Presented in Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Nursing Education

Nova Southeastern University

Richild A. Berrick
2019

NOVA SOUTHEASTERN UNIVERSITY
HEALTH PROFESSIONS DIVISION
RON & KATHY ASSAF COLLEGE OF NURSING

This dissertation, written by Richild A. Berrick under direction of her Dissertation Committee, and approved by all of its members, has been presented and accepted in partial fulfillment of requirements for the degree of

DOCTOR OF PHILOSOPHY IN NURSING EDUCATION

DISSERTATION COMMITTEE

_____   _____

Julia Aucoin, DNS, RN-BC, CNE   Date
Chairperson of Dissertation Committee

_____   _____

Erin E. Killingsworth, PhD, RN, CNE   Date
Dissertation Committee Member

_____   _____

Sarah Ransdell, PhD   Date
Dissertation Committee Member

NOVA SOUTHEASTERN UNIVERSITY
HEALTH PROFESSIONS DIVISION
RON & KATHY ASSAF COLLEGE OF NURSING

Certification

We hereby certify that this dissertation, submitted by Richild A. Berrick, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirement for the Doctor of Philosophy in Nursing Education degree.

Approved:

_____   _____

Stephanie La Manna PhD, MPH, ARNP, FNP- C, AGACNP-BC   Date
Associate Professor
Program Director PhD & DNP Programs
Ron and Kathy Assaf College of Nursing

_____   _____

Marcella M. Rutherford, PhD, MBA, MSN, RN   Date
Dean, Ron and Kathy Assaf College of Nursing

Abstract

Background: Effective testing in pre-licensure nursing programs is a challenge in nursing education. Implementing evidence-based test development is essential to successful assessment of students' competence and preparation for licensure.

Purpose: Identifying the beliefs, attitudes and values of nursing faculty will contribute to the use of best practices in student assessments, ultimately contributing to increased retention of competent students and increasing the workforce within the healthcare industry.

Theoretical Framework: This study is based on Rokeach's theory of beliefs, attitudes and values.

Methods: A quantitative descriptive research methodology was used in this study using survey data collection. A purposive, non-probability, convenience sample was the sampling strategy. The instrument utilized was developed and validated in a previous study and additional researcher-developed items were added. These additional items were field tested for readability and structure by current nursing educators.

Results: The results revealed that nursing faculty are not consistent with utilizing evidence-based test development practices within their nursing programs. The beliefs and attitudes identified from the data indicate a concern with the understanding and confidence towards evidence-based practices. Several challenges were identified in implementing test development practices such as addressing linguistic and cultural biases, faculty time constraints, and utilization of test banks.

Conclusions: Identifying faculty beliefs, attitudes, and values of evidence-based test development practices offers insight into the challenges facing nursing faculty, nursing

programs and nursing students. These challenges affect and influence the retention and

persistence of nursing students in prelicensure programs which ultimately affects

diversity in the nursing workforce.

Table of Contents

List of Tables

Chapter One

The Problem and Domain of Inquiry

The process of assessment is an integral component of nursing education and practice. Effective assessment of student learning and competence provides information about the quality of education provided. Nursing programs are accountable to their stakeholders and society regarding effective preparation of graduate nurses for their practice roles (Siegel, 2015). There are many assessment methods used in nursing education to acquire important information about student learning; however, the most frequently used strategy is the administration of examinations. Through effective testing, educators are able to make important decisions about students, courses, and curriculum. These tests are often weighted most heavily when calculating the students' grades in the course, which often become high stakes for the student.

According to the National League for Nursing's (NLN) Fair Testing Guidelines for Nursing Education (2012), there is an ethical obligation for nursing faculty to ensure that tests and all decisions made using information collected from these tests are valid and supported by sound evidence. Suskie (2009) explained that assessments need to be accurate, unbiased, and fair to all students. When an examination is biased, students perform differently based on variables that are unrelated to their knowledge and abilities. A biased test contains construct-irrelevant variances, such as item writing flaws, that may be confusing to students and can affect performance on the item. A test is fair when it is free of bias and students of equal ability are equally likely to answer it correctly

(Klisch, 1994). Therefore, it is imperative that nursing programs use unbiased, objective, and impartial assessments to afford all students an equal opportunity for success.

Nursing professional licensing standards require graduates to pass a comprehensive, primarily computer-adaptive examination. The National Council Licensure Examination for Registered Nurses (NCLEX-RN®) is not only used for licensure, but the scores are also used to determine accreditation status; it is used as the basis for program learning outcomes for all undergraduate nursing schools (Accreditation Commission for Education in Nursing, 2017). All state boards of nursing recognize the NCLEX-RN® as the exam that indicates whether newly graduate nurses have the minimal requisite knowledge and skills to provide patient care (Clarke, 2017). Currently, the National Council for State Boards of Nursing (NCSBN) determines what content will be included in the NCLEX-RN® and sets the nationally recognized passing standard (NCSBN, 2017). Most nurse educators are of the opinion that the best way to prepare students for the NCLEX-RN® is to mirror their course assessments with the licensure exam, which includes test format, content areas, and level difficulty. Because of the importance of effective preparation for nursing students to pass the NCLEX-RN®, increasing the quality of faculty-developed tests to accurately reflect knowledge gained and facilitate success on the licensure exam is a priority for stakeholders such as graduates, health care organizations, and nursing programs (Yeom, 2013).

Undergraduate pre-licensure nursing programs require nursing students to pass many high-stakes tests and assessments. Implementing standardized testing into the nursing curriculum has become popular in nursing programs to determine student progression and graduation. These commercial standardized tests are being utilized as a

form of high-stakes testing across the country. Academic progression and the prediction of passing the NCLEX-RN® are often dependent on the outcomes of these standardized tests. The use of standardized tests does not negate faculty responsibility to provide effective summative assessments throughout the course (Aucoin, 2008; Randolph, 2017; Richards & Stone, 2008). Teacher-developed tests are critical in the effective evaluation and preparation for these standardized assessments, the NCLEX-RN®, and clinical practice.

Effective assessment prepares nursing students for their professional practice. The application of critical thinking and nursing concepts in a dynamic healthcare industry is an expectation of graduate nurses. Management of complex clinical situations while providing competent care is the responsibility of the graduates of nursing programs (Morrison & Free, 2001). Tests that are designed to assess student outcomes related to providing safe and competent nursing care must, therefore, be written according to discipline-specific guidelines (Morrison & Free, 2001).

It has been discussed that the educational preparation of nurse educators has focused on clinical expertise rather than educational methodology and that many educators enter into academia without the necessary knowledge or expertise in teaching. Schmeiser and Welch (2006) also discussed the demand of educators to be vigilant in test design and focus on improving test and item development. Several authors have postulated that there are many variables that have impacted issues with effective test construction, some of which include the use of poorly constructed test banks, faculty inexperience, and lack of educational preparation. Based on these variables, misunderstandings and poor practices develop and flourish, and individual interpretations

of item writing guidelines in nursing examinations are common and persist (Bosher, 2003; Clifton & Schriner, 2010; Cross, 2000; Masters et al., 2001; Tarrant, Knierim, Hayes, Ware, 2006; Tarrant & Ware, 2008). Such violations often lead to poor student performance and poor retention in nursing programs, which, in turn, contribute to the nursing shortage. Therefore, effective test development practices by nursing faculty contribute to the successful preparation of students for success on the licensure examination and enhance quantity and diversity of the workforce.

Problem Statement

Developing reliable and valid assessments of student learning has long been an issue in nursing education. Faculty members often rely on test banks and other resources for their assessment tests (Nadeem, van der Vleuten, & Alfaris, 2012). Understanding and implementing effective test development in nursing education continue to be a challenge since most nurse educators were prepared as clinicians or managers, and not as teachers. Additionally, most nursing faculty, depending on their educational preparation, may have not received formal training in assessment practices and test development (Walloch, 2006). Therefore, they have minimal experience with test development, item writing, and other forms of assessment. Ultimately, effective testing plays an important role in curriculum and program planning, student placement, academic advancement, and educational research. If these results are influenced by nursing faculty's inability to effectively develop tests, then their validity is clearly reduced. There is little in the literature regarding how faculty perceive test development best practices and their ability to prepare the students with effective use of evidence-based methodology in test construction. Identifying the factors contributing to nursing faculty's beliefs and attitudes

motivating evidence-based test development practices will outline the variables impacting effective test construction and educational preparation.

Purpose of the Study

Very little empirical data are available concerning faculty beliefs regarding test construction in nursing education. Understanding teacher attitudes and beliefs about testing methodology issues may lead to a reduction in the frequency of questionable test development and improvement. Improving test development practices can ultimately increase the successful progression of nursing students within nursing programs and effectively prepare them for nursing practice. The purpose of this study was to describe the beliefs, attitudes, and values of nursing faculty regarding the utilization of test development best practices in undergraduate nursing programs.

Research Questions

This study aimed to uncover the beliefs, attitudes, and values of nursing faculty regarding evidence-based test development practices. As a result of the current challenges in test development processes in nursing education, the following research questions guided this investigation:

1. What features of test construction are utilized by nursing faculty in pre-licensure programs?

2. What features of test revision are utilized by nursing faculty in pre-licensure programs?

3. What are the differences in test development practices in pre-licensure programs between ASN and BSN faculty?

Significance of the Study

This study is especially significant to nursing students, nursing faculty, administrators of nursing programs, colleges and universities, state board of nursing administrators, and administrators of healthcare institutions. The data gathered have significant implications to add to the body of nursing by enhancing nursing education, practice, research, and public policy. An essential component of this study was to reinforce nursing faculty's accountability and responsibility in how to best assess NCLEX-RN® readiness and content mastery in a fair and effective manner, by ensuring that students are evaluated effectively on faculty-developed tests. This study highlights the current practices of nursing faculty regarding evidence-based test development and its effects on assessment processes.

Nursing Education

Providing additional research in evidence-based test development and faculty's beliefs and attitudes regarding its use offers insight into interventions to create a positive change in professional development about test construction practices. This research may ultimately contribute to increasing the retention of qualified students to meet academic expectations of graduate nurses in the community. Effectively describing the challenges faculty face in the utilization of best practices in test development and construction may lead to a more active stance on the use of these practices to enhance student assessments to achieve learning outcomes in undergraduate nursing programs.

Nursing Practice

The nursing shortage is a critical issue in the United States. The shortage is expected to get even worse since the demands for registered nurses (RN) are outstripping

the supply. According to the American Association of Colleges of Nursing (AACN, 2017), there will be more than 1.09 million new positions for RNs through 2024. Demands for RNs in acute care will increase by 36% through 2020, and a lack of more than 260,000 RNs will affect the U.S. healthcare system by 2025. Furthermore, because of the increase in the average age of RNs, it is expected that there will be an additional loss of a large number of RNs to retirement over the next 20 years. As a result of the shortage, nurses often need to work long hours under stressful conditions and maintain high patient-nurse ratios. These environments contribute to nurses' making mistakes and medical errors. In addition, insufficient staffing decreases patient safety, increases work-related stress and injuries for nurses, and causes nurses to leave the profession (AACN, 2011).

One of the strategies identified to resolve the problem of the nursing shortage is to expand nursing enrollment and retain nursing students while ensuring that they meet competencies to begin their nursing careers as entry level RNs. Retaining qualified students in nursing programs is important to meet the local and global need for registered nurses. Effective assessment practices will ultimately lead to retention of students and success on the NCLEX-RN®, leading to increased RNs in various practice environments. This study may provide a step forward in addressing the retention of nursing students and the evidence-based practices in educational assessment which may ultimately contribute to the increase in the nursing workforce.

Nursing Research

Research is a systematic process of investigation that uses disciplined methods to answer questions or solve problems with the main aim of expanding a body of knowledge

(Powers & Knapp, 1995). Nursing education research is designed to develop knowledge about issues of importance to faculty and students utilizing this methodological approach. Essentially, research enables nurse educators to describe the characteristics of a particular nursing education situation about which little is known (Polit & Beck, 2012). The development and utilization of nursing education knowledge are essential for continued improvement in student nurse preparation for practice. Nurse educators increasingly are expected to adopt research-based or evidence-based education practice using research findings to inform their decisions, actions, and interactions with students.

This study was conducted to address a critical gap in the nursing literature regarding faculty beliefs, values, and attitudes related to evidence-based test development practices in nursing education. Using a conceptual model with a contemporary social psychology focus, the study describes these beliefs and attitudes which greatly impact effective assessment of student learning and preparation for practice. This ultimately contributes to the body of nursing research by uncovering why the expectations of utilizing evidence-based test development in nursing education continue to not be met.

Public Policy

Lack of gender, ethnic, and generational diversity has been addressed by the National League for Nursing (NLN), focusing on the belief that the diverse patient population should have a nursing workforce that is as diverse in providing care (Phillips & Malone, 2014). Nursing programs are continually faced with the challenge of recruiting and retaining a diverse population. Diversity issues continue to affect many nursing programs across the nation. As the cultural landscapes change across the country, nursing schools face the challenge of recruiting and retaining students of various cultural

backgrounds. While today's nursing students are more diverse and nontraditional, testing practices in nursing programs have changed little in the past 20 years. Recently, nursing education experts have called for a radical change in student assessment methods (Benner, 2012; Benner, Sutphen, Leonard, & Day, 2010; Halstead, 2011; Kantor, 2010; Oermann & Gaberson, 2016). Researchers have clearly demonstrated that poorly constructed examinations penalize not only diverse and nontraditional students but all students and impact progression and retention in nursing programs (Bosher & Pharris, 2009; Carrick, 2011; Clifton & Schriner, 2010; Downing, 2005; Tarrant et al., 2006; Tarrant & Ware, 2012). Poorly constructed examinations continue to impact the success on the NCLEX-RN®. Poor outcomes on the NCLEX-RN® negatively affect the nursing programs, the nursing workforce, and the healthcare industry.

This study contributes to the understanding of the beliefs, values, and attitudes of nursing faculty related to evidence-based test development practices. This may improve the quality of teacher-developed tests and, ultimately, decrease the bias towards minority students and students who speak English as a second language, thereby creating equal opportunity for success and increasing diversity in the nursing workforce.

Philosophical Underpinnings

Paradigms, also considered worldviews, have a general philosophical orientation about the world and the nature of research that a researcher brings to a study. Worldviews arise based on discipline orientations and the types of beliefs held by individual researchers. Every research study must be grounded in the philosophical understandings about the world. Paradigms provide an empirical structure for exploring natural science to make sense of the nature of reality and the basis of knowledge (Creswell & Clark,

2007). In quantitative research, the post-positivist assumptions have represented the traditional form of research. This worldview is sometimes called the positivist/post-positivist research, the scientific method, empirical science, and post-positivism. Post-positivism represents the thinking after positivism, challenging the traditional notion of the absolute truth of knowledge and recognizing that we cannot be positive about our claims of knowledge when studying the behavior and action of humans (Budd, 2001). The post-positivist traditions come from 19th-century writers such as Comte, Mill, Durkheim, Newton, and Locke (as cited in Smith, 1983) and, more recently, from writers such as Phillips and Burbules (2000).

Epistemology (the nature of knowledge), ontology (the nature of reality), and the philosophy of science are paradigms of underpinning philosophical assumptions that influence the practice of research (Samdahl, 1999). The post-positivist epistemology is one of *modified dualism/objectivism*. This is the acceptance that all discovery is subject to interpretation; independence is not possible, but objectivity is seen as the goal during the discovery process. External validity demonstrates this objectivity. The ontological stance of post-positivism is *critical realism*. Critical realism is the belief that all observation is fallible and has error, and all theory can be revised. It posits that cause and effect relationships do exist, but it is not always possible to understand these relationships entirely. This ontological position believes that the research can be used to promote change and the researcher must be careful of value contaminating research (Mertens, 2014).

The most important assumptions of post-positivism that apply to this quantitative research study are as follows:

1. Data, evidence, and rational considerations shape knowledge in this research study. Researchers collect information on instruments based on measures completed by the participants or by observations recorded by the researcher.

2. This research seeks to develop relevant, true statements that serve to explain a situation of concern or describe the causal relationship of interest.

3. Subjective internal states such as feelings, attitudes, perceptions, and beliefs can be operationalized and objictified by translating them into numerical data.

4. Objectivity is an essential aspect of competent inquiry within this study. Methods and conclusions of this research study must be examined for bias such as reliability and validity.

Research methods can be traced back, through methodology and epistemology, to an ontological position. It is impossible to engage in any form of research without committing to ontological and epistemological positions. The differing ontological and epistemological positions of researchers often lead to different research approaches towards the same phenomenon. The quantitative methodology and post-positivist paradigm were applied to this research study (Scotland, 2012).

Theoretical Framework

The purpose of a theoretical framework is to organize the variables and relationships between the variables. The researcher can visualize the aspects of the study and the ways in which each variable will be measured and will interrelate (Morse & Niehaus, 2009). Selecting the appropriate theoretical frameworks is important to guide resolution of issues that may be identified, develop the research design, and establish a perspective for interpretation in order to suggest recommendations for the improvement

in nursing education. The framework for this study is based upon Rokeach's (1968) theory of beliefs, attitudes, and values.

The concepts of beliefs, attitudes, and values have been examined by many theorists. The beliefs, attitudes, and values theory by Rokeach contains constructs that are important and applicable to this research study. This theory is a cognitive consistency theory, which indicates that a person's beliefs, attitudes, and values must be in harmony with each other and which further explicates what the importance of this is while trying to persuade an individual. This theory utilizes a psychological perspective which examines how an individual's psychological characteristics affect and directly influence his or her behavior (Sereno, 2012). Its constructs are applicable to this study as they propose a solution to the attitude-behavior discrepancy issue. All three components of this theory are interconnected and will ultimately determine the behavior of the individual.

The first component of the theory is beliefs which are described as perceptions that something exists or that it is true. Beliefs are further classified as central beliefs and peripheral beliefs. Rokeach (1968) suggests that central beliefs are harder to change because of the centrality of their existence within the individual and are of more importance in determining behavior. Peripheral beliefs have lesser importance and are easier to change. In this particular study, the beliefs of the faculty related to evidence-based test construction will be influential in determining their potential or actual adherence to the best practices.

Attitude is the second construct of the theory. Attitudes are described as positive or negative feelings towards people, ideas, or things (Sereno, 2012). The important

element of this concept as it relates to this study is whether attitudes predict behavior and the relationship between the attitudes of the nursing faculty and their adherence to the best practices required to effectively construct valid and reliable tests.

The final concept within the theory is an individual's values. Values are considered beliefs that are important to a person that help guide their lives and are the most important construct of this theory. Values are the most important concept in this theory since by affecting a person's values, true persuasion occurs because these are beliefs that a person aims to live by. By changing the values of a person, all the attitudes displayed as a result of these closely held values must change; therefore, a number of the beliefs that are related to the attitudes displayed also change. Ultimately, the three variables involved—beliefs, attitudes, and values—form a functionally integrated system where a change in one part will affect all the other parts (Eckhardt, 1970). In understanding the values of the nursing faculty, the relevance and importance of best practices will result in determination of the use of best practices.

A study conducted by Clark-Goff and Eslami (2016) using Rokeach's theoretical framework focused on teachers' beliefs about English language learning and teaching. The study identified the beliefs and attitudes that teachers brought to the classroom. These beliefs were strong predictors of their behavior towards students who are English language learners (ELL). Because of the diversity of the student population and the high demands of high-stakes testing, the study demonstrated the importance of understanding mainstream teacher beliefs and how these beliefs can be altered. The study utilized surveys to collect the quantitative data from 354 participants, some of whom were then solicited for qualitative data through semi-structured focus group interviews. The

participants selected for the qualitative component of the study were determined by the

survey data obtained. The study found after implementing coursework addressing

diversity and English language teaching methodology, the teachers showed greater

acceptance of children and parents with native languages. Understanding the impact of

those beliefs was critical in effectively teaching this diverse population.

Another study utilizing the Rokeach's theory of beliefs, attitudes, and values was

conducted on neonatal nurses and their perceptions and choices based on their values

related to caring for certain types of patients. Raines (1994) utilized Rokeach's

theoretical framework to identify values influencing the nurses' behaviors in hypothetical

clinical situations. This descriptive study utilized a random sample of 331 members of the

National Association of Neonatal Nurses. The instrument was a researcher-developed

survey which focused on identifying the members' values based on three hypothetical

case scenarios. Results of this survey indicated that nurses had a hierarchy of beliefs and

values that guided their practice.

Rokeach's belief system is a conceptual model with a very simple premise:

Human beings have differing beliefs, attitudes, and values of differing intensity and

complex connections that determine their importance and behavior. These constructs

rooted in Rokeach's theory are important in the educational research of faculty beliefs,

attitudes, and values of evidence-based test construction practices. Beliefs, attitudes, and

values are all organized together to form a functionally integrated cognitive system, so

that a change in any part of the system will affect other parts, and will culminate in

behavioral change (Rokeach, 1968).

Assumptions

According to Rokeach's beliefs, attitudes, and values theory, one's belief system, including the substructures of values and attitudes, is a disposition to action and major determinants of behavior. Based on the assumptions of this theoretical framework, beliefs teachers hold will influence their perception and judgments, which in turn, will affect their behavior in their role as educators. All teachers hold beliefs, however defined and labeled, about their work, their students, their subject matter, and their roles and responsibilities. It is also assumed that understanding the belief structures of nursing faculty is essential to improving their professional practices.

Another assumption based on Rokeach's theory is if the belief system regarding test development is identified then it gives nursing education a point of reference to begin to understand the attitude-behavior discrepancy related to evidence-based practice. Identifying the attitudes and values of nursing faculty related to the professional task of test development can determine methods to address the issue. The connection between what nurse educators believe and the performance of certain educational tasks is the basis of this investigation.

There are three major assumptions made by the researcher regarding this study. The first assumption is that the faculty participants would answer the survey truthfully. The nurse educators are participating voluntarily and would be assured of anonymity and confidentiality. The faculty participants could also withdraw from the study at any time without ramifications. No participants withdrew from the study. The second assumption is that there would be a valuable number of responses to the survey disseminated to nursing faculty. In a similar study completed by Killingsworth (2013), the response was

sufficient and produced rich data in order to validate the findings. The final assumption

by this researcher was that the responses would provide a significant range of diverse

beliefs, attitudes, and values as it relates to the study constructs. These responses make a

considerable contribution to the research problem identified and to the body of

knowledge in nursing education.

Definition of Terms

The following terms discussed in this section are directly related to this research

and are used consistently throughout this document. These definitions allow for more

concrete understanding of the study and add to its relevance.

Beliefs

Theoretical definition: a cognitive act or state in which a proposition is taken to

be true. A belief is the state of mind in which a person thinks something to be the case,

with or without there being empirical evidence to prove with factual certainty.

Philosophers use the term to refer to personal attitudes associated with true or false ideas

and concepts that do not require active introspection and circumspection (Nespor, 1987).

Values

Theoretical definition: elements of an individual's belief system that measure

choices, whether consciously or not. Values are used to rationalize behavior of self or

others. They emerge from a combination of one's background, experiences, and evolving

sense of self (Rokeach, 1968).

Attitudes

Theoretical definition: a psychological construct, a mental and emotional entity

that inheres in or characterizes a person. Attitudes are complex and are an acquired state

through one's experiences. Attitudes are an individual's predisposed state of mind regarding value and are precipitated through a responsive expression toward a person, place, thing, event, or character (Richardson, 1996).

Pre-Licensure Nursing Program

Theoretical definition: according to the American Nurses Association, an educational pathway that culminates in one's being able to sit for the NCLEX-RN® licensure examination (Schmidt & MacWilliams, 2011).

Operational definition: Associate degree in nursing (ADN) is a two-year degree offered by community colleges and hospital-based schools of nursing in Florida. Bachelor of science in nursing (BSN) is a four-year degree offered at colleges, universities, and the few remaining diploma programs in Florida.

Chapter Summary

Retaining qualified students in nursing programs is important to meet the local and global need for registered nurses. Increasing the culturally and ethnically diverse nursing workforce is of paramount importance to meeting the needs of the current healthcare population. Healthcare organization leaders surmise that communication among healthcare professionals, quality of patient care, and overall safety will be enhanced by more diversity in the nursing profession. However, the retention of ethnically diverse nursing students remains low in nursing schools nationwide. The increased attrition of nursing students has been attributed to issues of bias in assessment practices (Andrews & Roberts, 2003; Bosher, 2003; Hicks, 2011).

Successful progression and graduation of nursing students are highly impacted by students' ability to meet the benchmarks of many high-stakes tests, including course

exams and standardized tests, within the nursing curricula. It is an obligation of nursing faculty to ensure that tests are fair, valid, and reliable to effectively assess student learning. The majority of nurse educators have minimal formal education or training in teaching and testing strategies. As a result, many assessment tests administered in nursing schools are poorly constructed (Clifton & Schriner, 2010; Cross, 2000; Masters et al., 2001).

Test development best practices and evidence-based guidelines must be utilized in assessment preparation in order to give all students a fair and equal opportunity for successful progression and graduation. Effective test development of quality nursing examinations also prepares nursing students for the NCLEX-RN® which ultimately leads to the increase in the quality, quantity, and diversity of registered nurses (Clifton & Schriner, 2010). Understanding the beliefs, attitudes, and values of nursing faculty regarding evidence-based test development is key to increasing the implementation of effective testing practices in nursing programs.

This chapter presented a discussion of the current challenges in the implementation of evidence-based practice in test development within nursing education; provided an overview of the effects on culturally diverse students; identified nursing workforce issues; explored the retention, progression, and graduation issues within nursing education programs; and established the impact of high-stakes testing in nursing curricula. The following chapter provides a thorough discussion of the theoretical and empirical literature related to this research study.

Chapter Two

Review of Literature

The development of valid and reliable tests in nursing education has been a continuing challenge. Faculty-developed tests can be erroneous, biased, and ineffective, a factor which significantly hinders the progression status of nursing students throughout their nursing programs (Dibattista & Kurzawa, 2011). This study focused on uncovering the beliefs, values, and attitudes that educators place on using evidence-based test development practices to effectively create tests that will assess students' competence and prepare them for clinical practice.

The most popular form of summative assessment in nursing education is testing (Popham, 2009). These classroom tests vary from teacher-developed to commercially-developed for purchase. Although classroom testing in nursing education is a dominant assessment technique, the focus on evidence-based test development practices has received scant attention in the nursing literature. Nurse educators have a professional responsibility and obligation to ensure that their test development practices are valid and reliable measures of student learning. It is the understanding of the beliefs, values, and attitudes of nursing faculty which may contribute to improving the testing practices within nursing education programs. A review of the literature will be presented in this chapter to provide information over five sections that correspond to the concepts and issues in current testing practices in nursing education.

Keywords, Databases, and Resources

The literature review included the following key words: test development, test items, multiple choice, test questions, nursing education, evidence-based nursing education, evidence-based practice, test bias, cultural bias, standardized testing, test construction guidelines, testing best practices, testing in nursing, assessments in nursing, faculty attitudes in test development, and faculty beliefs. The primary electronic databases used in the search were: the Cumulative Index to Nursing and Applied Health Literature (CINAHL); EBSCOHost; Education Full Text; ProQuest Nursing and Allied Health; ProQuest Dissertations; Educational Resources Information Center (ERIC); Academic Search Premier; and Education Research Complete. References frequently cited in relevant articles were reviewed, as were works by the key expert authors who developed the theoretical frameworks and taxonomies of test development guidelines. Other resources included Nova Southeastern University's library. The review is limited to approximately 75 English language articles.

The following section will comprehensively focus on literature supporting current best practices in test development. The discussion will also include current testing methodology and taxonomy, standardized testing, the NCLEX-RN® licensure examination, issues in test development, and identified linguistic and cultural biases.

Evidence Based Practice in Nursing Education

The Institute of Medicine (IOM) cited evidence-based practice (EBP) as one of the five essential core competencies for healthcare educational programs. According to the IOM, using EBP means the integration of the best research with clinical expertise and patient values to achieve optimal health outcomes (Long, 2003). Regardless of discipline,

EBP is a core competency for all healthcare clinicians. Nurses, specifically, are required to adhere to accepted standards of practice and professional performance. These standards mandate the use of EBP interventions and integration of findings into clinical practice. EBP requires nurses to access and appraise evidence before integrating it into practice (Mehrdad, Joolaee, Joulaee, & Bahrani, 2012). The focus of EBP has routinely been on the clinical care of patients and its integration into effective care.

Within the previous decade, EBP has also infiltrated nursing education. In academia, the focus of EBP has primarily been from the perspective of establishing active research within nursing courses and stimulating a research culture among nursing students. Currently, both baccalaureate and master's level nursing education programs have shifted the paradigm from an historical focus on preparing nurses who are generators of research to evidence users who effectively translate research findings into clinical practice (Fiset, Graham, & Davies, 2017). The American Association of Colleges of Nursing (AACN), the association that guides baccalaureate and graduate nursing education curricula, identified *scholarship for evidence-based practice* as an essential of baccalaureate education. A priority for nurse educators has been to focus on developing positive attitudes towards EBP by role modeling and developing the necessary skills to integrate evidence into nursing courses (AACN, 2008).

The use of EBP to effectively prepare nurses for the dynamic healthcare environment is a priority for nursing programs. Shifting the focus from teaching EBP to implementing it in nursing education practices is critical. Integrating empirical research in instructional strategies, student assessment processes, and curricular development increases the retention and success of nursing students nationwide (Fiset, Graham, &

Davies, 2017). According to the National League for Nursing (NLN), the academic nurse educator must achieve identified competencies to promote excellence in this role. There are expectations of the complex faculty role as a specialty area of practice. One of the competencies is the use of assessment and evaluation strategies. This competency describes the responsibility of the nurse educator to "use a variety of strategies to assess and evaluate student learning in classroom, laboratory and clinical settings, as well as domains of learning" (NLN, 2005, p. 3).

Increasing importance has been placed on improving student learning and retention in nursing education programs; however, there is little evidence about whether nurse educators are aware of using research and how much research based strategies are being used in the academic environment (Whorley, 2018). Understanding the attitudes and barriers towards the integration of research by nursing faculty can potentially influence the effectiveness of methodology. This can ultimately have a positive effect on the preparation of the nursing workforce. Utilizing EBP in nursing education is a way to improve nursing education practices and student outcomes. In meeting these practices, educators can ensure that graduates are fully prepared to meet the workplace requirements which is critical to the nursing profession and, most importantly, to the safety and optimal health of patients. Testing in nursing supports the identification and reporting of student progress, evaluating student learning, identifying individual and group challenges in learning, and providing important data regarding curriculum and course efficacy (Oermann & Gaberson, 2016). The types of testing in nursing education significantly impact nursing students' success and progression (Poindexter, Hagler, &

Lindell, 2015). Implementing EBP related to the types of testing in nursing education is a vital component of preparation for professional practice.

Testing in Nursing Education

A significant amount of attention in nursing education is focused on classroom assessment. It is customary for nurse educators to use faculty-developed tests to assess the progress of nursing students in the didactic portion of the nursing programs. A major focus of the testing practices is to effectively prepare students to pass the NCLEX-RN® and successfully achieve licensure. More importantly, it is the professional responsibility of nursing faculty to prepare students for safe and effective practice. In addition to licensure and safe practice, nursing programs must meet accreditation standards related to student and program outcomes (Siegel, 2015). The Commission on Collegiate Nursing Education (CCNE) is an autonomous accrediting agency which strives to ensure the quality and integrity of baccalaureate, graduate, and residency programs in nursing. Recognized by the United States Secretary of Education as a national accrediting agency, the CCNE sets forth accreditation criteria related to standards which include the mission and governance, institutional commitment and resources, curriculum and teaching-learning practices, and assessment and achievement of baccalaureate program outcomes. Specific criteria related to the effectiveness of the nursing program and its outcomes are the program's NCLEX-RN® pass rate and their nursing student completion rates, which is the outcome of the program's assessment and testing practices (CCNE, 2017). The Accreditation Commission for Education in Nursing (ACEN) is another accrediting agency which provides designated accreditation for different types of nursing transition to practice education programs. ACEN's role supports the interests of all nursing

programs (e.g., diploma, ASN, BSN, MSN) and continually monitors these programs through state and licensing rules. The agency also provides oversight of the preparation of students by nursing programs for transition into the nursing profession (ACEN, 2017).

The National League for Nursing (NLN) has also established core competencies for the academic leader in nursing education. Specifically, core competency number three focuses on the use of assessment and evaluation strategies in nursing programs. The NLN expects nurse educators to "use a variety of strategies to assess and evaluate student learning in classroom, laboratory and clinical settings, as well as all domains of learning" (NLN, 2012a, para. 5). The NLN also offers the certified nurse educator (CNE) certification in order to "distinguish academic nursing education as a specialty area of practice and an advanced practice role within professional nursing" (NLN, 2012a, para. 1). A major content area of the CNE examination is the focus on effectively using assessment and evaluation strategies. Incorporating current research in assessment and evaluation practice and creating effective assessment instruments to evaluate outcomes are significant skills nursing faculty must demonstrate (NLN, 2011).

Based on the NLN requirements for nurse educators and the CCNE and the ACEN's program accreditation criteria, the centrality of nursing assessments and testing in nursing programs significantly contributes to the success and efficacy of a nursing program. It is this reality that reinforces the need for EBP in test development and assessment practices in nursing education. The following sections will focus on the types of testing in nursing education, which include course testing, standardized testing, and preparation of the NCLEX-RN® licensure examination.

Course Testing

Whereas nursing courses remain testing oriented, an increasing number of publications discussing testing techniques recognize the implications on effective student assessment (Aucoin, 2005; Clifton & Schriner, 2010; Duty, Christian, Loftus, & Zappi, 2016; Siegel, 2015; Su, Osisek, Montgomery, & Pellar, 2009; Tarrant & Ware, 2008). Faculty-developed assessment tests are the primary method of student assessment in undergraduate nursing programs. In many instances, it is the sole method of didactic assessment. This is the reason the faculty-developed tests must be well constructed to assess higher cognitive levels and critical thinking (Siegel, 2015). Nursing faculty must be efficient in understanding the complexity of assessment in the classroom and ultimately create effective assessment tests. However, most nursing faculty have not received formal training in test development and effective assessment construction within their educational preparation programs (Walloch, 2006).

Assessment and instruction methods are interrelated and are impacted by the educator's beliefs and values of the process. Effective assessment is significantly affected by how nursing faculty plan and deliver their instruction. The incongruence between the assessment and the teaching methods continues to raise concerns (Kantar, 2014). A teaching methodology is often misaligned with testing practices, which ultimately affects the student assessment data obtained. Nursing faculty continue to be saturated with course content and instruct at a knowledge cognitive level, but they implement higher cognitive level testing with their students. This leads to a disconnect between instruction and assessment and errors in the validation of students' ability to meet course objectives (Oermann & Gaberson, 2016).

An analysis was completed by Clifton and Schriner (2010) in a baccalaureate nursing program involving the faculty-developed final examinations. There were a total of 256 tests selected for evaluation. The researchers randomly selected 20 items from each test for analysis. The intention of the faculty was to assess the students' ability to apply and analyze concepts; however, the test items selected were evaluated at a knowledge and comprehension cognitive level. In addition, 63% of the number of items analyzed were labeled as being extremely below the cognitive level necessary for that course. Most of the conclusions were based on the poorly written items and the weak and implausible distractors within the question. The subject of implausibility of distractors in test development has a significant effect on the validity and reliability of the examination (Rodriguez, 2005). The use of textbook test-banks is another concern in test development for course examinations. Faculty members often rely on the course textbook test-bank in order to develop their course exams. Research conducted by Masters and colleagues (2001) focused on the review of nursing textbook test-banks. The study focused on analyzing the test-banks for their adherence to item development guidelines by Bloom's taxonomy of cognitive levels (1956) and Haladyna and Downing's (1989) multiple choice question guidelines. The authors analyzed approximately 3,000 items and discovered that over 2,000 violated item development best practices. Most of the items were categorized as low level knowledge questions which are not effective in assessing and preparing nursing students for the NCLEX-RN® examination. No more recent studies were found in the literature regarding test bank analysis. Textbook test-bank items are seldom validated for proper psychometric properties and often contain many biases and best practice errors. Many of the items are also labeled with an incorrect cognitive level.

Faculty's heavy reliance on test-banks to develop nursing tests can cause invalid assessment of nursing competence and impact progression and retention of nursing students (Downing, 2005; Tarrant, Knierim, Hayes, & Ware, 2006; Tarrant & Ware, 2011).

According to Notar, Zuelke, Wilson, and Yunker (2004), faculty-developed tests have historically been incompetently designed as compared to the reliability and validity of items on the NCLEX-RN® and other standardized achievement exams. In a study by Cross (2000), 110 instructor-developed final exams were randomly selected from 66 nursing programs in the United States. The analysis revealed that over half of the items developed by the faculty for the course exam were poorly written and erroneously leveled based on Bloom's taxonomy. In a dissertation by Killingsworth (2013), a descriptive correlational design was used to analyze faculty's use of best practices in test construction. The researcher investigated faculty in baccalaureate nursing programs in over 30 states. The data revealed that many educators do not always use best practices in test construction. Participants reported using course objectives, class objectives, and specific objectives to develop their tests. The data also revealed that peer review and the use of the NCLEX-RN® test plan for test development were used the least. Assessment of linguistic and cultural bias was the least used in the test revision process, with item analysis based on difficulty being the most used practice. Many of the course test construction practices reported by the participants did not include blueprinting, item analyses, peer review, use of current EBP, and identification of linguistic bias consistently.

The impact of fallacious test items has been discussed substantially in nursing education literature. The impact of poor course testing and its effects on student success must be a priority of academic leaders. Faulty test development practices penalize all students and have a major impact on the effective preparation for clinical practice. Tarrant and Ware (2008) determined that issues with poorly developed test items in nursing courses had a negative impact on the higher achieving students more than the borderline students. Bosher and Pharris (2009) and Clifton and Shriner (2010) also discuss the impact of ineffective test development practices on diverse and nontraditional students. Progression and retention of these students are significantly influenced by substandard assessment practices. Improving the quality of faculty-developed tests is a topic that must be addressed by administrators, educators, and nursing organizations.

Standardized Testing

Standardized national tests have become very popular within nursing education programs as a predictor of NCLEX-RN® performance. Nursing programs integrate these standardized tests to protect and predict their NCLEX-RN® pass rates (Randolph, 2017; Richards & Stone, 2008). However, the effectiveness of these standardized assessments continues to stir up controversy. In many nursing programs these standardized tests are considered high stakes by requiring the nursing students to achieve a minimum score in order to progress or graduate. Some of the more popular standardized testing exit exams are Health Education Systems Incorporated (HESI™) HESI E-2, Assessment Technologies Institute (ATI™) Comprehensive Predictor Exam, Kaplan Readiness Test, NLN End of Program, and Mosby Assess Test. Meeting the benchmark for these assessments can determine if, ultimately, the student will be denied the ability to graduate

and sit for the NCLEX-RN® licensure exam (Nilbert & Morrison, 2013). In addition to the exit exams, the predictive precision of the standardized course content tests are also utilized throughout nursing curricula. Standardized course content tests are utilized at the end of individual courses in order to assess the students' comparability nationally. These course tests are often also assigned an assessment (grade) percentage for the class, thereby determining the progression of the student from term to term (Richards & Stone, 2008).

In 2012, the NLN responded to the increasing concern of the significant reliance nursing programs had on standardized testing as a graduation requirement. The organization developed the NLN Vision: Fair Testing Imperative in Nursing Education and NLN Fair Testing Guidelines in order to address the arguable use of standardized high-stakes testing. The NLN's Chief Executive Officer, Dr. Beverly Malone, expressed concern regarding eligibility denial of students to taking the NCLEX-RN® based solely on standardized exams and deemed it unfair to students and possibly harmful to the nursing profession (NLN, 2012b). The NLN offered recommendations for test development and implementation based on evidence on testing practices from nursing and other disciplines. Although the NLN suggests multiple sources and approaches for assessment of knowledge and clinical abilities critical to evaluating students' abilities in addition to high-stakes testing, many nursing programs continue to utilize standardized testing in decision making regarding progression and graduation (NLN, 2012b).

Preparing students effectively is crucial to the performance on national standardized tests. A survey completed by the NLN on assessment and grading practices in nursing programs established that the most important factor guiding faculty decisions

on strategies to use in assessment of student learning outcomes is the NCLEX-RN[®] pass rates (NLN, 2012b). To emulate the NCLEX-RN[®] exam, nursing faculty believe that utilizing multiple choice questions on course exams is the most effective form of assessment. With such *high-stakes* testing being administered in nursing programs, it is imperative that these tests contain quality question items that are reliable and valid as a means to assess student learning. A general guideline of the NLN Fair Testing Guideline for Nursing Education describes the ethical responsibility of the nursing faculty to establish validity of their tests as well as the decisions based on these tests. The NLN further emphasizes the need for substantial evidence and consistency to ensure fairness for all test takers within their programs (NLN, 2012b).

Effective test development in nursing education is imperative for the preparation of students for standardized tests and the NCLEX-RN[®]. Quality items are essential in identifying strengths and weaknesses in students and their ability to meet course and program objectives. However, quality items are difficult to write, and poorly developed test items are often misinterpreted and fail to assess the students' abilities effectively. The construction of reliable and valid exams challenges even the most experienced nurse educators (Nadeem et al., 2012). According to DiBattista and Kurzawa (2011), most college faculty have very little formal education or training in teaching and testing strategies, and this fact also significantly applies to nurse educators. Graduate nursing programs provide minimal coursework on student assessment and effective item writing, so nursing faculty are often ill-prepared to successfully navigate the test development process, thus contributing to the challenge of poorly constructed tests in pre-licensure nursing programs. Essentially, the utilization of unreliable and invalid test items in test

development has a significant effect on the successful preparation of nursing students for standardized exams and the NCLEX-RN®.

NCLEX-RN®

The NCLEX-RN® tests the mastery of entry-level practice requirements for all pre-licensure nursing students. It is a standardized exam that each state board of nursing uses to determine whether a candidate is prepared for entry level nursing practice. The examination content and behaviors are differentiated by several areas considered *client needs*, including (a) management of care-20%; (b) safety and infection control-12%; (c) health promotion and maintenance-9%; (d) psychosocial integrity-9%; (e) basic care and comfort-9%; (f) pharmacology and parenteral therapies-15%; (g) reduction of risk potential-12%; and (h) physiological adaptation-14% (NCSBNa, 2016). Each of these areas within the NCLEX test plan can guide nurse educators to focus on curriculum development and student success.

The exam plan also consists of integrated processes which include the nursing process, caring, communication and documentation, teaching/learning, and culture and spirituality. The NCLEX-RN® exam items encompass all eight client needs categories and integrated processes and are based on a practice analysis survey distributed to entry level nurses. The practice analysis focuses on 142 job activities performed and the current practice of the entry level nurse. The results of the survey and the job activities identified guide content distribution of the client need categories on the NCLEX-RN®.

The passing scores are based on a minimum of 60 items to a maximum of 250 items, most of which are at the application or above cognitive level based on Bloom's taxonomy for the cognitive domain (Clifton & Schriner, 2010; NCSBN, 2016b). Fifteen

additional items are added to the candidates' exam as research items for future use,

bringing the total minimum number of items to 75 and the maximum number of items to

265 answered by the candidate. Item formats within the examination include multiple

choice, multiple response questions, *Hot Spot* items, fill-in-the blank calculation items,

and drag-and-drop/ordered response items. All item types may have additional audio

visual components such as charts, tables, graphics, sound, and video (NCSBN, 2016a;

Wendt & Kenny, 2009). The exam is administered through computer adaptive testing

which utilizes an algorithm to tailor the difficulty of each item to the candidate's ability.

The computer continuously evaluates the candidate's ability and selects items that he or

she should find challenging. The computer must be 95% confident in the candidate's

ability based on the number of items they have answered for the candidate to pass the

exam (Sanders, 2018).

First time NCLEX-RN® pass rates are used to measure nursing program quality

and curriculum effectiveness. Nurse educators have the responsibility and accountability

to effectively prepare students to pass the NCLEX-RN® as well as effectively prepare

safe and competent nurses. Essentially, nursing students who have successfully

completed a nursing degree should be qualified to achieve licensure through the NCLEX-

RN® examination which measures minimal competency to practice nursing. This minimal

competence should be achievable by utilizing valid and reliable assessment practices

throughout the nursing programs (Tarrant & Ware, 2008).

The NCSBN is the organization established to protect the public by ensuring that

all licensed nurses are competent to practice (NCSBN, 2016a). The NCSBN focuses on

developing the NCLEX-RN®, an exam which has both psychometric integrity and

defensibility and is continually updated. The updates are intended to ensure that entry level nursing practice is able to meet the dynamic healthcare environment. Prior to 1994, the NCLEX-RN® licensure exam was administered by paper and pencil biannually. Today, the NCSBN uses the computerized adaptive testing (CAT) method for the licensure exam. As stated previously, candidates receive a minimum of 75 questions to a maximum of 265 with a maximum time for testing of 6 hours. Once the computer has determined that the candidate, with 95% certainty, has scored either above or below the standard, the session will be ended. If the candidate fails the exam, the NCSBN stipulates that there must be at least 45 days between the original attempt and the retake (Carrick, 2011).

Each state board of nursing establishes its own rules and regulations regarding retesting. In Florida, for example, the candidate must take a review course if the candidate has failed the NCLEX-RN® three times. After six attempts, the candidate must complete an additional nursing program before being eligible to retest (Florida Board of Nursing, 2018). In a study by Griffiths, Papastrat, Czekanski, and Hagan (2004), 21 participants who had failed the NCLEX-RN® at least once were asked to identify the factors they believed contributed to their outcome. The factors identified were poor program preparation, inadequate study habits, lack of knowledge on how to prepare, lack of confidence, poor test-taking skills, overwhelming family responsibilities, and employment. In an additional study by Aucoin and Rodgers (2000), an additional factor identified was demanding scheduling factors faced by the students. There are multifaceted components to examine when nursing students who graduate from a nursing program and fail the NCLEX-RN®. Effective preparation and assessment of students can

minimize the risks of these students' failing the licensure exam. Nursing faculty must implement evidence-based assessment strategies and instructional methodology to prepare these students to pass the NCLEX-RN® on the first attempt.

Guidelines for Test Development

High-stakes consequences are often attached to tests in nursing education. These consequences affect students, course faculty, nursing programs, healthcare communities and institutions, accrediting agencies, and licensing boards (Clifton & Schriner, 2010). The purpose of these tests is to provide data from which conclusions regarding students' achievement of learning objectives and competence are determined. Effective test development plays a significant role in providing valid data to reach these conclusions. Researchers have gathered and outlined best practices for test development for which many have been validated through systematic research. This section of the study will delineate best practices and evidence related to research-based guidelines for test development.

Bloom's Taxonomy

A widely used framework which classifies statements based on the learning expectations of students is the taxonomy of educational objectives (Anderson & Krathwohl, 2001). Originally, the framework was designed as an avenue of encouraging the exchange of test items among faculty at various educational institutions in order to develop banks of test items, each measuring the same educational objective. This idea originated at the University of Chicago by Benjamin S. Bloom. As the associate director of the board of examinations, Bloom had the expectation to help decrease the intensive labor process of comprehensive examination preparation by university faculty. A number

of statistical specialists were assigned to this project, each from different universities across the United States. The official product of this collaboration was published in 1956. The title of the document was *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. This document was then known to educators as the original Bloom's taxonomy (Anderson & Krathwohl, 2001).

The original taxonomy has six major categories within the cognitive domain that were strategically and precisely defined in order to promote effective usage. These categories are knowledge, comprehension, application, analysis, synthesis, and evaluation. The structure of these categories went from simple to complex and from concrete to abstract. This ultimately led to the universal use in education to classify curricular objectives and test items (Krathwohl, 2010).

The knowledge category of Bloom's taxonomy is the fundamental cognitive skill and focuses on the retention of specific information. Examples of this type of information include facts, definitions, methodology, and sequence of events or processes. Effective assessment of knowledge can be completed by straightforward questions in multiple choice format or short answer questions that often require basic memory and retrieval of the information (Krathwohl, 2010).

The comprehension category of the taxonomy is the next higher level in the cognitive domain. Comprehension is focused on the students' showing they understand the meaning of the information and can explain the concepts. This category requires more cognitive processing than simply recalling and remembering information. Within this category, students can demonstrate comprehension by classifying items in groups,

comparing and contrasting concepts, and explaining principles to colleagues (Krathwohl, 2010). The next higher cognitive level is the ability to apply the information learned.

Application of a concept or principle involves the use of knowledge and skills to address new situations. Students may be asked to solve problems with the knowledge they have gained within their coursework. As the student moves through the cognitive levels, the next level is the analysis category of the taxonomy. It is within this category that students are often required to critically think about a given situation. The student must be able to utilize the information learned with the information given to break down a situation into its component parts in order to identify the most appropriate action. Identifying patterns and developing conclusions based on the analysis of the information are requirements in this category (Krathwohl, 2010). Following the analysis level is synthesis.

Synthesis focuses on utilizing the information provided to create new theories or make predictions in a given situation. Multiple sources of information must be used in order to synthesize information before reaching a conclusion. This involves a higher level of critical thinking in the student. The final and highest level of Bloom's taxonomy of cognitive domain is the evaluation level. The expectation at this level is that the student can assess information provided and come to a solid conclusion as to its value or bias. Evaluating situations based on previous knowledge and additional information provided can often be a challenge for students who do not have strong critical thinking and analytical skills (Krathwohl, 2010).

Forty-five years following the development of Bloom's taxonomy, the framework was revised as were the nomenclature and order of the cognitive processes by his

students. In this later version of the taxonomy, the levels were renamed to now include

remembering, understanding, applying, analyzing, evaluating, and creating. The

reorganization or revision of the taxonomy now placed the skill of synthesis (create) at

the highest level rather than evaluation as seen in the original taxonomy. The revision

went further to add new subcategories to the hierarchy. The knowledge category was

subcategorized into the following: factual (terminology and discrete facts); conceptual

(categories, theories, and principles); procedural (technique, process, or methodology;

and metacognitive (self-assessment ability). This restructured taxonomy, hereafter, is

referred to as the revised taxonomy (Anderson & Krathwohl, 2001).

Nurse educators focus on the development of the student's critical thinking and

clinical reasoning abilities. Developing learners' thinking skills at the higher level of the

taxonomy are necessary for students to demonstrate deeper cognition processing and

evaluation skills. Test construction and item development are often based on these

cognitive levels in order to assess if the learner has met the learning objectives. However,

it has been demonstrated in many research studies that nurse educators develop tests that

are far below the cognitive level required to effectively assess the student's status and

progress as well as his or her critical thinking abilities. It is crucial to the successful

preparation for student nurses to be tested at the appropriate level with examinations that

are valid and effective in educational assessment (Poindexter, Hagler, & Lindell, 2015).

Taxonomy Development

Successful test development is based on a systematic, precise, and detail-attentive

process that is based on substantial educational measurement principles. The data (test

scores) must provide sufficient and valid evidence to support the proposed inferences.

The terms test development and test construction are used to describe the many

components and details used to create an effective assessment tool. The test development

process must be well-organized to manage all the details necessary to produce a well-

prepared test that estimates the examinee's achievement or ability. This process must be

consistent in its ability to measure and support the test score inferences (Poindexter,

Hagler & Lindell, 2015). The most comprehensive evidence-based practice on test

development was conducted by Haladyna and Downing (1989) who developed a model

of systematic test development. This system is organized into 12 discrete tasks or

activities. This model was developed based on considerable research into this issue and

relevant processes discussed by the Standards for Educational and Psychological Testing

as they pertain to test development. These standards are based on information from the

American Educational Research Association (AERA), the American Psychological

Association (APA), and the National Council on Measurement in Education (NCME).

Each one of the 12 steps of test development is associated with one or more standards

(Downing & Haladyna, 2011). Table 1 lists the 12 steps of test development and provides

a brief summary of each task and activity.

Table 1

Twelve Steps for Effective Test Development (Downing and Haladyna 2011).

Steps	Test Development Tasks
1. Overall Plan	Prepare a systematic guide for all activities: purpose, format, construct, psychometric model, timeline, security control

2. Content definition	Purpose of assessment, sources of content-related validity evidence; delineation of construct
3. Test specifications	Framework for validity; norm or criterion referenced; desired item characteristics
4. Item development	Adherence to evidence-based principles; training of item writers; reviewers; effective item editing
5. Test design and assembly	Creating test forms; selecting items for forms; planned blueprint; pretesting considerations
6. Test production	Printing or packaging; security issues; timing issues
7. Test administration	Validity issue concerned with standardization; ADA issues; proctoring, security issues; timing issues
8. Scoring test responses	Validity issues; quality control; test key validation; item analysis
9. Passing scores	Establishing defensible passing scores; relative versus absolute; cut scores; comparability with standards; constancy of score scale
10. Reporting test results	Validity issues; accuracy; quality control; timely; meaningful; misuse issues; challenges; retakes
11. Item banking	Security issues; usefulness; flexibility; principles for effective item banking
12. Test technical report	Detailed documentation of validity evidence; recommendations

Note. This table was adapted from the original table by Downing and Haladyna (2011).

The first step in the test development process is constructing an overall plan for the test. This plan is a systematic framework to developing the test. This step focuses on the essential components associated with the effective development of tests such as the timeline, ensuring the security of the test, and controlling quality from the initial starting point. The critical decisions about the test being developed must be made prior to implementing any test development activities. Examples of these critical decisions include the purpose of the test, the content description and definition, the type of test, and the psychometric model to be used. Essentially, no matter the type and purpose of the test, effective and detailed design activities in the overall plan will be significantly beneficial (Downing, 2006).

The content definition of a test is the second step in the effective development of a test. The majority of achievement tests rely significantly on content-related validity evidence to ensure that arguments can firmly support the inferences of the test scores. The constructs and content being tested must be clear and definitive in order to be measured effectively. Content-defining methods need to be appropriate for the type of test being developed. Lower stakes tests may not have content definition criteria as stringent as those of the higher stakes tests. Higher stakes tests such as final exams, nursing exit exams, and the licensure exam must be methodical, organized, comprehensive, and defensible. This is especially vital when decisions and consequences are made based on the validity of the scores from these tests (Downing, 2006).

The third step in the development process is the test specifications process, also known as test blueprinting. This is the process of creating specific guiding details and activities that systematically complete the operational planning in test development. A

test blueprint is a specific outline of questions that assigns the number of questions to each area of content. These questions are also intended to assess different cognitive levels as described by Bloom's taxonomy or the revised Bloom's taxonomy. A test blueprint also outlines the types of question structure such as multiple-choice questions (MCQ), select all that apply (SATA), sequencing, and *hot spots*. Detailed test blueprinting provides significant validity evidence for the test data. Table 2 presents a simple example of a test blueprint for a medical/surgical course unit test (Downing, 2006).

Table 2 Sample Exam Blueprint

Content	Remember	Understand	Application	Analysis	Synthesis	Evaluation	Totals
Outcome	4	2	0	0	5	3	14
Outcome	2	4	10	10	3	1	30
Outcome	1	4	8	6	4	1	24
Outcome	2	1	5	5	0	3	16
Totals	9%	11%	23%	21%	12%	8%	100%

Step four in test development is the systematic development of selected-response items. Item development is a very precise process with well-established principles of item writing. The challenge during this step is to develop effective test questions that are created to measure the identified content at the appropriate cognitive level. The most widely used item structure is the MCQ because of its versatility and ability to test higher level cognitive functioning. Although the MCQ is the most versatile and popular, it is the most challenging to develop for item writers. Once developed, the clear and succinct rationale for each item is required. Evidence-based item writing has a solid base in the

research literature (Downing, 2006). Although there are clear guidelines and evidence-based practices for item writing, this does not guarantee the item writer will utilize these principles to develop effective items. Lack of training is often an issue with item development. Novice item writers often develop poorly constructed, flawed, low cognitive level, and biased test items. Developing well-written items leads to valid evidence for high-stakes testing. According to Downing (2004), some of the worst test items are found on instructor-developed tests at all levels and areas of study. This has been described as one of the major failings of educational assessment at all levels of education.

The fifth step in the test development process is design and assembly. The focus during this step is the quality control of the test. Depending on the mode of administration (e.g., paper-pencil or computer), formatting and effective assembly are key to the successful administration of the test and the data it produces. Some of these formatting issues are related to the readability of the item, location on the page, visual or graphical stimuli, and placement of correct answers in the items. Unless the quality controls and security of the tests are observed, errors will decrease the validity and legitimacy of the final test scores and skew the interpretations (Downing, 2006).

Step six is the production of the test. This step is often underestimated as it pertains to the validity of the data collected from the test. This step includes the printing, packaging, and publishing in some form. One of the major issues in test production is test security. More than one individual can often access final drafts of an exam. This access can be during any time prior to administration of the test. All security protocols and policies must be adhered to in order to secure the test. Printing of tests in a common area

is also a challenge in the test production step. Most faculty share printers within their departments and often encounter printing issues that may cause security breaches during test production. This breach in security of test production can also alter the final scores and validity of the test (Downing, 2006).

The seventh step of the test development process is the actual test administration. The conditions under which a test is administered can also affect the validity of the test results. Proper proctoring, environment, and time limits must be adhered to when administering a test, especially high-stakes tests found in healthcare education programs such as nursing. Proctoring is by far the most important component of this process as it is a major concern in administration of tests. Ensuring that proctors are oriented and trained to effectively maintain security during test administration is essential to obtaining valid test scores. Deficiencies in this step can lead to a significant reduction in the validity of the evidence of the examination. Security issues can also lead to the elimination of valid high quality test items and test forms, invalidation of test scores, and score misinterpretation because of the compromise of the test administration procedure (Downing, 2006).

The proper scoring of the test responses is the eighth step in the test development process. Creating measurement of a test is based on the examinees' responses to the test items. This creates fundamental validity issues with scoring of tests. In order to have validity to the test, the scoring (answer) key must be utilized with extreme accuracy to the item responses. Errors in scoring tests will notably reduce the validity evidence of the test and may often invalidate the results. Key validation is a process utilized to analyze the item and perform a preliminary scoring of the test data. This allows the test developer to

identify potentially flawed or incorrect items before the final scoring of the test. This two-stepped process is vital for tests containing newly developed and non-tested items. These items are especially vulnerable to invalidating flaws that were not discovered during the item writing and review process (Downing, 2006).

Another major step in this process of test development is establishing the passing scores for the test. The methodology used to determine these passing scores has a direct implication on the validity of the test evidence. In this step nine, the following key questions need to be asked when determining what the passing score should be: how much knowledge is needed to have a passing score and what skills or abilities must the examinee demonstrate in order to be considered proficient. The use of standard setting methods such as relative standard setting and absolute passing score methods can be used to meet the requirements of this vital step in the process (Downing, 2006).

Reporting the examination results is the tenth step of test development. This step is often one of the more complex and important steps in the process. The precision of this step is in line with many of the previous steps discussed, and attentive and effective measures to control quality of the test is of significant importance. High-stakes exams are specifically important to the efficient and accurate exam reporting results because of the effect on the examinees' status of pass or fail. Examinees are entitled to the accurate reporting of their test scores without deviation or over-interpretation. The method and language used to report the scores to the examinees should be congruent with their level of understanding. The accuracy and readability of the test scores are essential to the testing process and increase the legitimacy and the validity of the test results reported (Downing, 2006).

Development of effective test items is time intensive and requires specific skills in order to meet the test development standards. The secure storage of these test items is significant to the testing process. The difficulty in developing test items necessitates adequate resources to secure these items and their supporting data for future use without compromise. This eleventh step focuses on the item banking of effective test items to ensure their security and ability to be utilized in future versions of the test. The cost and labor needed to develop test questions warrant the secure item banking in a convenient, efficient manner with easy retrieval to assist test developers in increasing the validity evidence of the examination (Downing, 2006).

The final step in this process is the test technical report, which documents the test development process, the administration, scoring, reporting, and analysis of the test. This step is the final component that serves to provide the effective documentation of all the validity aspects of the entire test. The report can also identify areas that need to be reevaluated and recommendations needed for improvement. Technical reports also preserve essential validity for archival record and can significantly contribute to the testing development process (Downing, 2006).

The twelve-step test development process as presented by Haladyna and Downing (1985, 1989) provides the structure and systematic process needed to effectively produce valid and reliable tests in any academic program. Every test developed in an academic program should include most of the steps presented. It is imperative that the higher the stakes of the test, the more strict and structured the adherence to this process. The precise attention to detail in all areas of this process will more validly measure the objectives and domains of interest of the examinee. The validity and accuracy of the test development

process significantly contribute to effective testing throughout academia, including nursing education.

CRESST Criteria

The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) is an organization located at the University of California, Los Angeles (UCLA) that strives to improve and elevate student learning and assessment to the highest level internationally. CRESST focuses on evidence-based research in assessment design, implementation, scoring, and high quality measures as they relate to diverse educational environments. The nucleus of the organization is to improve evidence-based inferences related to psychometric assessment models that can be utilized at any level of education worldwide. CRESST has developed criteria for effective test development that encompass seven areas. The areas are as follows: cognitive complexity, content quality, meaningfulness, language appropriateness, transfer and generalizability, fairness, and reliability (Baker, Aschbacher, Niemi, & Sato, 1992).

Cognitive complexity identifies the various levels of learning that can be assessed. With the goals of the instruction being the reflection of the test, the test questions must focus on the appropriate intellectual activity required. This ranges from simple recall to critical thinking and reasoning. This cognitive complexity is based on the previously discussed original taxonomy of educational objectives that includes knowledge, comprehension, application, analysis, synthesis, and evaluation. Depending on the objective of the test being developed, it is imperative to formulate the structure of the test and utilize effective cognitive complexity to measure the appropriate level of student knowledge (Baker et al., 1992; CRESST, 2018).

Content quality is the next CRESST criterion for test development. Content quality in essence is the ability of the assessment to measure what it is intended to measure and for its intended purpose. There are three main levels to content quality: assessment design, assessment experience, and the assessment item level. The assessment design is structured by the content blueprint, which clearly delineates the content that will be assessed on the test and the cognitive rigor of that particular content. The assessment experience is based on the type of testing the student will experience, whether it is a traditional fixed form paper/pencil test or a computer adaptive test which provides students with questions based on their level and ability to answer. The CAT testing is a higher quality assessment tool, as it will provide more specific data for each student's learning and performance in the areas identified. The assessment item is also key to the validity of the assessment test and the level of knowledge being tested. There are several ways that content quality can be supported. These include general assessment design principles that control for readability, evidence-centered design methodology, and statistical analysis of student performance on test items. Overall, the content quality is foundational to making accurate inferences about the learning and what it is meant to measure (Baker et al., 1992; CRESST, 2018).

The third criterion of CRESST is the meaningfulness of the test. Meaningfulness of an assessment is simply the value of the test to the student. The items and focus of the test should be valued and be worth the student's time. Students should recognize and understand the values of the assessment they are being given. The items should focus on skills and knowledge that are necessary for the future of that student. This creates

meaningfulness for the student, which ultimately increases the validity of the outcomes (Baker et al., 1992; CRESST, 2018).

Language appropriateness is an essential component of the CRESST criteria for test development. Test items should be clear and language simple and free of nonfunctional material or extraneous clues. Effective test development involves the use of items that are free of racial, ethnic, and sexual bias. In today's diverse educational environment, students' culture and language backgrounds can have a major impact on their test outcomes. Focusing on the vocabulary and the syntax of the test may alleviate language barriers. The language demands must be clear and appropriate to the assessment tasks and to the students. In addressing language appropriateness, test developers can create effective assessments that provide valid measurements of learning (Baker et al., 1992; CRESST, 2018).

Transfer and generalizability of tests focus on the ability to utilize the successful performances on tests to validate generalizations about achievements by students. The data from an effectively designed and developed test can be used in generalizations to influence instructional placements, diagnostic decisions related to the course, and formative evaluation and also allow faculty to understand the needs of the student. Ultimately, this can lead to effective changes in courses, curricula and instructional design (Baker et al., 1992; CRESST, 2018).

CRESST criterion six focuses on the standard of fairness of the test. Tests must be developed based on the objectives of the material covered, and the expectations must be clearly understood by the students. The items must also be graded constructively and fairly with appropriate feedback that helps the student understand what they have

achieved by taking the test. Allotting the appropriate time for the student to take the test is also a component of fairness in test development. This will ensure that students have enough time to answer the questions to the best of their ability which ultimately will lead to valid and reliable data (Baker et al., 1992; CRESST, 2018).

The final criterion of effective test development is focused on the reliability of the test. The reliable data accrued from the test will demonstrate the level of knowledge and comprehension of the students and will be consistent and trusted. The reliability of a test will enable the faculty to make key decisions and judgments regarding students' successful attainment of the objectives and goals required for that particular course competency. If a test is unreliable, the data obtained will be useless in the educational process (Baker et al., 1992; CRESST, 2018).

Test development is an important process as it measures significant learning in a way that supports desired performance. Educators must utilize the evidence-based processes to effectively measure deep understanding and student learning. The CRESST criteria focus on sound research-based methods and criteria to develop effectual testing processes to achieve valid measurements of learning. Faculty in all areas of academia who develop tests must utilize current performance assessment research in order to meet the technical quality of the assessment. CRESST research continues to focus on the goal of effective assessment of performance for reliable and valid data of student learning which ultimately is used to make judgments and determinations about students' progression and success (Baker et al., 1992).

Key Principles in Test Development in Nursing

Sutherland, Schwartz, and Dickison (2012) developed four principles of best practice in test development. The principles are based on an accumulation of research related to evidence-based test development practices. Sutherland et al. (2012) synthesized this literature and provided the four principles for test development as an evidence-based guide for faculty to utilize in developing, assessing, and utilizing classroom tests to assess competency in nursing. The four principles are alignment, importance, differentiation, and fairness.

Alignment is the first principle established to ensure successful development of nursing tests. Defining and identifying the appropriate body of knowledge, skills, and abilities, also referred to as the *domain*, for the test is paramount in an effective process. A test blueprint must be created identifying tasks, activities, course objectives, and key concepts to be assessed for competency. The alignment of the test is how precise the test follows the established blueprint. A test that is not fully aligned with the stated objectives or goals cannot adequately determine the examinee's competency and/or proficiency in those areas. When tests consist of items that are inadequately aligned, no clear and concise conclusions can be made about any of the examinees. This includes both the high performers as well as low performers. Based on this significant reliability issue, alignment of the test must be a critical component in test development (Sutherland et al., 2012).

An effective test must be able to measure concepts in meaningful ways. This is the focus of principle number two which is the *importance* of the test. As discussed in much research related to nursing content overload and/or saturation, it is evident that

course testing cannot test students on everything taught. So it is essential that the objectives being tested are important and meaningful in determining proficiency and competency of the student. The informal *so what?* rule can be applied to the importance of items on a test. If an item reviewer or testing committee can identify items on the test and say "so what?" this should lead to the idea that the item is not important or meaningful in determining the competency of the student (Sutherland et al., 2012).

The third principle is that of *differentiation*. Providing substantial information about a student is the purpose of testing in nursing education. This information is intended to differentiate students by their ability and performance. Tests that have items that do not differentiate examinees provide useless data. Items that are answered correctly by all students and/or items that are answered incorrectly by all students do not measure what was intended. In order to effectively accomplish the differentiation, faculty test developers must identify the cognitive level of the test item and the distractibility of the incorrect answers. Focusing on the effective use of both testing components can improve the differentiation and efficacy of a test. Therefore, differentiation of item levels and difficulty is an imperative process in test development (Sutherland et all, 2012).

Fairness is the final principle in test development. The appropriateness and the applicability to all students testing are essential. This fairness principle focuses on the linguistic and cultural diversity of students and the nuances of test items that may inhibit their ability to effectively demonstrate competency. In test development, faculty must pay close attention to consciously avoiding linguistic and cultural bias that is unrelated to the student's knowledge, skills, or abilities. Lack of fairness of a test will also decrease the validity and the reliability of the results. Outliers also affect the fairness of the test.

Outliers are specific answer choices that are very distinct in their correctness or incorrectness that the student automatically is distanced from being considered. Outliers affect the fairness of the test based on the student's using test-taking strategies to respond instead of their knowledge, skills, and abilities (Sutherland et al., 2012).

Applying these four principles to the test development process in classroom faculty-made tests is essential to effectively assessing students' competency. Proper training of nursing faculty in test development and item writing is significant for measuring what is intended to be measured. Providing resources to ensure these principles are adhered to will enhance the test development process and impart reliable and valid data with which decisions of student progression can be made.

Another critical element to valid test development is identifying biases in testing. This issue mainly includes cultural bias and linguistic bias. This major component of test development will be address in the following section.

Bias in Testing

Diversity in the profession of nursing has been an issue of concern for many years. Increasing the diversity of nurses to reflect the community they serve has been an important initiative for the profession (AACN, 2015; Ayoola, 2013; Bednarz, Schim, & Doorenbos, 2010). However, the attrition rates of racial-ethnic minority students in nursing programs continue to be disproportionate to those of their white colleagues. One of the main contributing factors discussed in the research has been the issue of biased testing practices in nursing education (Bosher & Pharris, 2009; Lampe & Tsaouse, 2010; Bednarz et al., 2010). Cultural and linguistic biases have been areas of concern as they

relate to the fairness of assessments and testing practices. In this section of the study, the testing biases in nursing education are discussed.

Cultural Bias

Testing in nursing education intends to measure competency and proficiency with nursing concepts. Most of these tests are norm based on concepts and values of majority groups, which can be biased against minority groups including race, gender, language, socioeconomic status, and culture (Hicks, 2011). Performance gaps between ethnic minority groups and non-minority groups still remain apparent despite interventions to decrease them. Nursing education has seen an increasing amount of culturally diverse students applying and entering nursing programs nationally. This diversity ultimately will have a positive impact on the nursing profession and improve care for the diverse community served. However, the success rates for students from many cultural groups have challenges with succeeding in nursing education programs primarily in their ability to pass faculty-made tests. Much of this issue is based on the test development process and the identification of cultural bias within the test itself (Amaro, Abriam-Yago, & Yoder, 2006).

Cultural bias essentially refers to the content in test items that is not equally available to all cultural groups. Understanding special or specific cultural knowledge in order to answer a question on a test is considered cultural bias as it contains references to a particular culture, customarily the majority group (Bosher & Pharris, 2009). Culturally biased terminology used in test items usually are not essential to demonstrating safe and effective nursing practice. Klisch (1994) completed a study within a nursing program by reviewing all the nursing examinations developed by faculty. It was identified that all

tests developed in that program had culturally biased terminology and inferences. Masters et al. (2001) evaluated a 2,913 item random sample from test banks developed by textbook publishers and found that 2,223 items were flawed with extensive cultural bias as a significant component. It is imperative that nursing faculty evaluate items found in test banks for flaws, especially related to cultural bias. In addition, Tarrant and Ware (2008) discuss the lack of time for evidence-based test construction and formal preparation in assessment methods that contributes to the epidemic of cultural bias in nursing education tests.

Multicultural students continue to have difficulty with testing in nursing education. In addition to meeting the program learning outcomes and objectives, these students must also learn the language and the U.S. culture (Bednarz et al., 2010; Bosher & Bowles, 2008; Bosher & Pharris, 2009; Hicks, 2011). Most of the areas with which these students struggle are related to therapeutic communication, spousal and child abuse, child-rearing practices, and nutritional and dietary preparation. Humor (jokes and puns), terminology, and stereotyping are also typically areas in which cultural bias is evident (Lujan, 2007). Addressing cultural bias in test development has ethical implications because of the disproportionate number of multicultural students failing out of nursing programs (Bosher, 2003; Bosher & Bowles, 2008; Hicks, 2011; Lampe & Tsaouse, 2010). Considering that most testing in nursing education is based on faculty-made tests, it is crucial that the development of these tests be free from bias in order to afford equal opportunity for success to these students.

Linguistic Bias

An increasing number of non-native speakers of English have enrolled in nursing programs nationally. These students often have difficulty in succeeding academically due primarily to the language difficulties. These difficulties are predominantly seen in testing within the programs (Lampe & Tsaouse, 2010; Taxis, 2002). Linguistic bias on nursing exams is a significant challenge to students with English as an additional language (EAL). Most faculty in nursing programs today continue to utilize test banks in order to develop their classroom tests. Linguistic bias continues to plague questions developed by publishers of textbooks and resource websites. It is the responsibility of nursing educators to analyze nursing program tests in order to minimize linguistic bias in test development (Camilli & Shepard, 1994).

In a qualitative study completed by Dudas (2014), participants discussed the profound effect of language on influencing their nursing education experience. The participants expressed the challenges they faced and that linguistic issues played a major part in achieving academic success. The participants discussed the need for extra time on tests, reading assignments, and presentations. In Dudas' personal account of her experiences with EAL students, she discussed the attrition rate of a university where she was a faculty member. Within this university, 76% of the students in the baccalaureate program were minority EAL students. Of the students ultimately dismissed from the university, nearly 80% were students who spoke English as a second language (Dudas, 2014). In the study by Killingsworth (2013), the findings identified that assessing for linguistic or cultural bias in test items was not a priority for the participants in implementing best practices in test construction and revision.

The concept of linguistic bias is essentially the use of unnecessary linguistic complexity within a test question that contributes to lack of clarity and consistency in the wording. This also leads to the use of construct irrelevant variance (CIV) which ultimately leads to skewed outcomes. CIV is the process of introducing superfluous, uncontrolled variables that assess the test's outcomes. These variables lead to test results that contain invalid data and adversely affect the testing legitimacy and the decisions derived from them. These CIV are more likely to produce false results of students' competency and proficiency with the subject matter being tested (Hicks, 2011; Lampe & Tsaouse, 2010).

Some of the issues in test development that contribute to linguistic bias include irrelevant difficulty, embedded and reduced clauses, and unclear wording. The irrelevant difficulty describes the test items that intentionally incorporate content that is not relevant to the focus of the item. This challenges EAL students to spend more time on the question trying to process the unrelated information in the item instead of demonstrating their actual understanding of the content (Klisch, 1994). Test items also have embedded and reduced clauses in the item. This occurs when test developers embed clauses in the questions that force the student to read and reread the question several times to understand what is being asked. Unclear wording is a very common form of linguistic bias. This occurs when test developers utilize wording that is not common terminology. Unable to clearly understand the meaning of the uncommon term, EAL students often attempt to interpret the word in order to demonstrate understanding of the concept. This often leads to an incorrect answer based on wording and not on actual competency (Abedi, 2006; NCSBN, 2006).

According to Klisch (1994), linguistic and structural bias in testing consists of test items that are intentionally "long, unclear, or contain awkward or misused grammatical construction[s], these items contribute to noise rather than information, to the testing process" (p. 36). According to Bosher (2009), the NCSBN considers an item free of bias if "the probability of responding correctly, given total score, is the same for all subpopulations" (p. 46). Addressing the linguistic bias of nursing examinations must be a commitment for all nursing education to improve the success rate of multilinguistic and international students in nursing programs. Effective and fair evaluation processes begin with a fair and unbiased test development process.

Nursing faculty are often unprepared to develop valid and unbiased tests within the courses they teach. A significant contributing factor is the lack of formal or informal training. Nurse educators must be adequately prepared to meet the expectations of the role. The issues identified with nurse faculty preparation for their educator role will be discussed in the following section.

Nursing Faculty Preparation in Test Development

Nurse faculty are often employed in nursing programs based on their clinical expertise rather than their educational preparation and experience. The transition from expert clinician to novice educator is often a challenging one (Benner, 2012). New nurse educators often underestimate the skill set required for an effective nurse educator and struggle with the faculty role and all its components. Effective orientation and mentoring of new faculty are essential to the adequate transition to the educator role. Most orientation and mentoring programs focus on effective teaching approaches, student engagement, policies and procedures, and classroom management (Anderson, 2009). A

significant responsibility during the transition process often overlooked is the evidence-based test development process. In addition, seasoned faculty often lack the skills needed to meet this critical component of nursing education orientation. Resources are often provided for faculty to develop nursing tests such as textbook test banks and NCLEX-RN® preparation books. Nurse educators are also encouraged to create their own test questions without the proper training necessary to create fair and reliable tests. This leads to the challenges often identified in nursing programs related to effective assessment and progression and retention of nursing students (Clifton & Schiner, 2010; Killingsworth, 2013; Oermann & Gaberson, 2016). There are many professional development programs focusing on test and item development for nurse educators by nursing education organizations; however, faculty often focus on other skills for nursing education practice.

Educational Preparation of Nurse Educators

Nurse educators are employed in a variety of nursing programs and educational settings. NCLEX preparation occurs at the pre-licensure level of nursing education, including associate degree, diploma, and baccalaureate programs. This level of nursing education requires specific knowledge and skill competencies as it relates to effective preparation for licensure and transition into clinical practice. Nurse educators are also required to meet certain core competencies as stated by the NLN as well as advanced educational preparation. The role and responsibilities of these educators include the ability to effectively assess student competency with valid and reliable methodology. Test development skills are a key component of effective assessment in nursing programs.

In today's educational climate, most nurse educators enter into academia as clinicians with very little preparation as nursing faculty. The formal educational

preparation of these educators employed in post-secondary settings is determined by the state boards of nursing and programmatic accrediting organizations. The accrediting agencies vary in their requirement standards of formal preparation of nursing faculty. The CCNE, for example, requires the nurse educators to be academically (graduate degree) and experientially prepared for the areas they teach (CCNE, 2018); however, the ACEN does not specify the degrees required. The NLN Commission for Nursing Education (CNEA) requires the nurse educators to be credentialed by their education preparation, professional experience, and teaching responsibilities (Oermann & Gaberson, 2016). Some boards of nursing require preparation as an educator through academic preparation, continuing education, or certification. However, most post-secondary institutions lean more towards the requirement that faculty members hold a graduate degree in nursing in order to transition into the academic environment.

Many graduate nursing programs have an education track for nurses aspiring to enter into the academic setting. These tracks are based on the NLN's *The Scope of Practice for Academic Nurse Educators* (2012). Found exclusively in doctor of philosophy (PhD) in nursing education, master's, and post-master's certificate programs, these tracks focus on the theoretical and practicum components of nursing education. In addition to formal education, the NLN recommends that nurse educators continually engage in professional development activities that will contribute to the quality of education for nursing students. The framework developed by the NLN for graduate programs has eight domains with specific competencies required for the nurse educator role. The competency which encompasses test development skill is the ability to develop assessment and evaluation strategies (NLN, 2005).

In 2005, the NLN issued a position statement entitled *Transforming Nursing Education* in which they discussed the importance of nursing education's being research-based and that appropriate faculty preparation for the academic role includes demonstration of multiple competencies. Along with this call for competence, the academic nurse educator role became recognized as a practice specialty (NLN, 2012a). This led to the NLN's establishment of the academic nurse educator certification program (ANECP). The ANECP then developed the Practice Analysis Committee (PAC) with members who represented diversity within the profession to identify job-related responsibilities of individuals who would fulfill the full range of the nurse educator role. The committee, based on the data obtained from the practice analysis, organized the eight core competencies of the certified nurse educator (CNE®) credential (Ortelli, 2006). The use of assessment and evaluation strategies within the faculty role is one of the core competencies of the CNE, which specifies the testing component of the knowledge, skills, and abilities that characterize this professional role. Research shows increased empowerment, knowledge, expertise in meeting practice standards, professional commitment, and credibility for nurse educators who obtain the CNE credential (Nick, Sharts-Hopko, & Leners, 2013).

Assessment experts assert the importance of effective student assessment in meeting learning and program outcomes (Nadeem et al., 2012; Oermann & Gaberson, 2016; Poindexter et al., 2015; Popham, 2009; Siegel, 2015; Sutherland et al., 2012). Unfortunately, the longstanding issue in nursing education regarding constructing valid and reliable tests in nursing programs continues to be a threat to nursing programs and nursing students. Identifying faculty beliefs, attitudes, and values related to effective test

development using best practices may contribute to developing effective interventions to address this continued challenge.

Gaps in the Literature

Formal preparation of nurse educators in test development practices through graduate programs is important to the transition into practice. However, it is difficult to obtain specific information related to test development preparation strategies within these programs. Most research studies focus on other aspects of nursing education competencies such as instructional methodology, scholarship, clinical education practices, and simulation. There is scant research focused on the specific amount and level of faculty education and competence with test development in graduate nursing programs in order to prepare faculty for the academic role. Essentially, faculty often transition into nursing education with little or no formal training in student assessment and test development. Improving the competence of nurse educators in test development is significant to effectively determining competence of nursing students within nursing programs. This study serves to add to the body of nursing education research by including this critical component of nursing education. By identifying the beliefs, attitudes, and values of nurse educators related to evidence-based test development practices, the researcher focused on this gap in the nursing education literature.

Chapter Summary

Competent test development practices are a significant component of nursing education programs. Effective testing and assessment will produce reliable and valid data regarding student competency. Flawed test development processes and poor faculty-developed tests interfere with the accuracy and meaningful interpretation of exam scores

and can have a negative effect on student progression in the program. The failing of high-stakes exams in nursing programs is often due to poorly developed tests. Many of these tests contain poor test development practices, cultural and linguistic bias, and exaggerated use of test banks. These exam scores provide invalid data that are utilized, often exclusively, to make decisions on students' academic standing and possibility of graduating. There are significant implications of nursing students' failing poorly constructed tests because of the difficulties instructors have with developing valid and reliable tests. The need for implementation of evidence-based test development practices is paramount in addressing this issue and will ultimately improve the quality of nursing education and testing practices. This dissertation study may contribute to the nursing education body of knowledge by identifying the beliefs, attitudes, and values of faculty regarding the use of evidence-based practices in test development presented. Identifying these beliefs can contribute to improved test development practices for all levels of testing in nursing programs. This chapter presented a thorough review of the theoretical and empirical literature related to evidence-based test development practices and methods of improving test development practices in nursing education programs.

Chapter Three

Methodology

The primary purpose of this dissertation study is to describe the beliefs, attitudes, and values of nursing faculty regarding best practices in test development and their impact on the utilization in their assessment practices. In addition, this study will describe the findings that may contribute to future studies of evidence-based test development in nursing education.

Research Design

This study is broad in focus in order to gather data on a large scale with a sizeable number of nurse educators. A sample size of 278 usable surveys was anticipated. This sample size was based on the Modified Cochran formula for small populations with a 95% confidence level and a 0.5 plus or minus deviation. The final sample size for this study was 117. The data collected were self-reported about faculty members' beliefs, attitudes, and values about evidence-based practice in test development. This study implemented a quantitative, descriptive design. Quantitative research designs promote objective, rigorous, and systematic strategies for generating and refining knowledge. Quantitative research designs are classified as either experimental or nonexperimental (Sax, Gilmartin, & Bryant, 2003). The non-experimental design of this study is descriptive.

A quantitative research methodology is appropriate for this study since a survey was used for data collection. The purpose of the survey research is to generalize from a

sample population so that inferences can be made about some characteristic, attitude, or behavior of this population. It also identifies attributes of a large population from a small group of individuals (Punch, 2005). By using this method, identification of faculty best practices across several variables and statistical comparisons between groups can be made (Dillman, 2000). Advantages of survey design are the economy of design and the rapid turnover in data collection. This methodology is used to enrich the description of a particular phenomenon in nursing education about which very little is known (Polit & Beck, 2008). The qualitative approach was excluded because it did not align with the purpose of the study which was to produce statistical data. This study was not seeking to develop a theory, interpret cultural constructs, focus on ethnography, discuss lived experiences, or describe themes as seen in phenomenology (Polit & Beck, 2008). Utilizing the highly structured methodology of quantitative design, the researcher intended to produce objective statistical data for the basis for future studies.

Descriptive studies have an important role in educational research and have greatly increased our knowledge about what happens in the academic environment. Descriptive designs identify challenges that actually exist, in which there is little or no previous research completed (Sax et al., 2003). It was the intent of this descriptive research to present statistical information about aspects of evidence-based test development, an important component of nursing education, which interests all faculty and student stakeholders. This descriptive study yielded rich data that will lead to important recommendations regarding assessment practices in nursing education.

Research Assumptions

The research assumptions of this study included the following:

- Best practices exist for test development.

- The methodology is appropriate to answer the research questions.

- Beliefs, attitudes, and values influence nursing faculty behaviors.

- The research instrument is valid and actually measures the variables under investigation.

- The participants will meet the eligibility requirements.

- The participants will answer the survey truthfully.

- The data analysis techniques are appropriate to interpret the data obtained.

- The data collection process will not influence the participants or the results of the study.

Setting

There are currently 123 associate degree in nursing (ASN) pre-licensure programs offered and 44 bachelor of science (BSN) programs offered in Florida. All pre-licensure program graduates are expected to satisfactorily complete the NCLEX-RN®. The setting for this study was pre-licensure nursing programs in the state of Florida with the target population being any nursing faculty currently participating in the preparation of examinations within the nursing program.

Sampling Plan

The principal goal of sampling is to obtain a representation of a larger population to contribute as sources of primary data. In obtaining this smaller population in Florida, the researcher can study and produce accurate generalizations. The most current NCLEX-RN® performance in Florida reveals passing rates from 33% to 98% with an average of 84.24% in all pre-licensure programs (Florida Board of Nursing, 2018). Effective

sampling establishes accuracy of the research findings and has tremendous implications on the overall quality of the study. Yielding highly representative samples requires specific processes, which are significant to the validity and the generalizability of the study (Creswell, 2014). The approach and type of sampling will be outlined in this section of the study.

Sampling Strategy

A purposive, non-probability convenience sample was used in this study. The non-probability, purposive sampling format was used to direct the survey to the deans and directors of nursing programs in Florida. To facilitate access to the purposive sample, the deans and directors were asked to forward the email-linked survey to the faculty within their programs who are involved in test development practices to access the appropriate participants. Currently, there are 167 pre-licensure programs in Florida. To minimize the sampling error for the survey, 117 classroom faculty at undergraduate, pre-licensure schools of nursing in the state of Florida were recruited.

The initial step in the sampling process was compiling a list of the deans, directors, or chairs of each of the nursing programs through the Florida Board of Nursing (FBON) website. These nursing leaders were contacted to seek their support in forwarding the survey to their nursing faculty using an email invitation (see Appendix A). The participant letter and consent form were attached to the email (see Appendix B). The majority of the leadership contact information was obtained from the National League for Nursing (NLN) and American Association of Colleges of Nursing (AACN) databases. In addition to the nursing leaders, the email invitation was directly sent to nursing faculty

known to the nurse researcher. Two follow-up emails were sent to the deans and directors on week two (2) and four (4) of the open survey period.

Eligibility Criteria

Recruitment of the participants included nursing faculty currently teaching and engaged in test development practices in pre-licensure nursing programs in Florida. Eligible to participate in this study were those faculty who participate in item writing and test construction within their nursing programs. Participants were excluded if they were not actively engaged in test development practices within their nursing program.

Inclusion criteria

Inclusion criteria are characteristics that prospective participants must have in order to be included in a research study (Creswell, 2014). The inclusion criteria for the participant selection in this study included all nursing faculty who actively participate in test development and item writing within their Florida pre-licensure nursing program.

Exclusion criteria

Exclusion criteria are considered those characteristics that disqualify participants from being included in the study (Creswell, 2014). The exclusion criteria for this study included all nursing faculty who were not actively involved in test development processes within their Florida pre-licensure nursing programs, such as those who function only as clinical setting instructors.

Determination of Sample Size: Power Analysis

Based on the geography and the inclusion criteria, the survey was sent to 110 pre-licensure nursing programs in Florida. The response rate anticipated was 20-25% of the surveys sent to potential participants. This percentage is the customary response rate for

electronic surveys (Hart, Brennan, Sym, & Larson, 2009). As stated by Polit and Beck (2008), an effective size of 0.50, power of 0.80, and an alpha of 0.05 are appropriate for nursing research studies. An important component of the research studies is the effective size as it focuses on the degree to which the independent variable(s) influences the dependent variable (Pallant, 2010). Significant statistical findings are ideal; however, effect size will demonstrate the strength of the relationship between the independent and dependent variable (Field, 2013). Based on the Modified Cochran formula for smaller populations, the suggested sample size was to be approximately 278 participant surveys for this study. The actual number of surveys analyzed was 117.

Protection of Human Subjects

The ethical considerations and protection of human subjects consisted of several steps. First, the researcher submitted to the institutional review board (IRB) of Nova Southeastern University for permission to conduct the study. Second, after receiving permission from IRB to conduct the study (see Appendix C), an email invitation was sent to the deans, directors, or chairs of the nursing programs to introduce the study. The email invitation included a thorough description of the purpose of the study, procedure, risks, benefits, and confidentiality. In addition, the email was sent directly to potential participants. The privacy and confidentiality of all the participants were maintained at all times.

Risk and Benefits of Participation

Risks to the participants in this study were minimal and related to the time involved in completing the surveys and any difficulty encountered with technical issues. Every effort was taken to minimize the respondent burden during this data collection

phase. The benefit to the participants in this study was enhancement of their self-awareness as it relates to their test development practices in nursing education.

Data Storage

A structured data collection and storage process is important in protecting research participants. It is essential to include best practice methodology and procedures to protect the privacy and confidentiality of participants. It is the duty of the researcher to ensure that participants' privacy and confidentiality is consistent throughout the research study (Punch, 2005). The research data was collected and managed using the Research Electronic Data Capture (REDCap$^®$), an electronic data capture tools hosted at Nova Southeastern University. REDCap$^®$ is a secure, web-based application designed to support data capture for research studies, providing: 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources (Harris et al., 2009).The primary investigator and the dissertation committee were the only ones to have access to the data. The emails to multiple participants employed the blind distribution option in order to protect the privacy of the email addresses. All email account data were erased at the end of the participant recruitment. All data were reported in aggregate form, and no identifying information was included in any written report of the research.

<div align="center">

Procedures

</div>

Data were collected using a survey instrument that was accessible through a survey link distributed via email by the REDCap$^®$ platform. Potential respondents were invited to participate in the study by an email message that contained an embedded

hyperlink to a website hosting the instrument; the research relied upon a gatekeeper, the dean or director leader of the program, to forward the survey. The online data collection was selected for several reasons: (1) the research population has access to email accounts and high speed Internet connections from their employing universities; (2) participating research sites can be surveyed at the same time period, providing for consistency in survey timing and administration; (3) the immediacy of survey invitations and the ease of completing the survey were expected to produce a good response rate from research participants; and (4) electronic submission of surveys was provided for immediate transfer of data responses into a database that can be exported into a spreadsheet for analysis purposes (Fricker & Schonlau, 2002).

The survey delivery platform REDCap® was implemented once the study had been approved and permissions granted by the institutional review board (IRB). Once the faculty clicked on the first eligibility link, the survey opened (see Appendix D). At the end of the last screen with the survey questions, the respondent was thanked for his or her participation.

Instrumentation

The descriptive function of research is heavily dependent on instrumentation for measurement and observation. Once instruments are developed, they can be used to describe phenomena of interest to the researchers (Polit & Beck, 2008). The Best Practices in Test Development instrument created by Killingsworth (2013) has demonstrated validity and reliability. The primary investigator in this study obtained permission (see Appendix E) to utilize the tool with authorization for additional questions as needed for this study.

Research Instrument

The survey instrument utilized in this research study addresses several components of the test construction and decision-making processes used by nursing faculty. The instrument aimed to identify the beliefs, attitudes, and values of nursing faculty in using evidence-based practices in test construction. The survey is based on an original tool developed by Killingsworth (2013) with additional sections of items added by this author related to beliefs, attitudes, and values of the participants. The first section of the survey focused on the demographics and teaching background of the participants. There are 13 of the 16 original items from the original survey by Killingsworth (2013). This section addressed credentialing, experience in teaching, current nursing program employed in, and standard demographic information.

The second section of the survey focused on test construction practices of the participants. The participants were instructed to indicate the frequency of each component used during their test development process. The aim of this section of the survey was to identify what practices and how often these practices were being used in the test development process. The Likert-type scoring from *not at all* to *all the time* is utilized within this section. The higher the score in this section, the higher the use of evidence-based test development practices. The participants were also instructed to identify sources of test items that were used during test development. This sub-section is scored from *never use* to *always use*.

The third section of the instrument focused on the test analysis and revision process. This component of the instrument aimed at obtaining data related to the best practices the participants used to analyze tests after administration and how their test

revision processes were completed following the analysis. The Likert-type scale within this section was based on how often these practices are used by indicating *not at all* or *all the time*. The higher the score in this section, the higher the use of best practices in test analysis and revision by the participants.

The fourth section of the instrument focused on teacher beliefs regarding assessment practices in nursing education. These items aimed to obtain information regarding the participants' personal beliefs related to general evaluation of students and what is best for students' classroom setting assessments. The participants were instructed to indicate on the scale of *least descriptive* to *most descriptive* how the evaluation belief applies to them. Based on the scoring of this section, the higher the score the more concerned the participant is about fair and effective evaluation methods and their effect on student assessment.

The next section focused on the participant's beliefs, attitudes, and values regarding evidence-based test development practices. The 24 items in this section, developed by this author, aimed to understand the perspectives of the participant in utilizing evidence-based practices in test development. These items were based on Rokeach's theory of beliefs, attitudes, and values as well as information obtained in an extensive literature review. The items in this section were field tested for readability and usability by 15 nursing faculty in a pre-licensure nursing program. This section is scored from *strongly disagree* to *strongly agree* related to how the practice statements describe their beliefs, attitudes, and values. Based on the scoring of this section, the higher the score, the stronger the beliefs, attitudes, and values are placed on using evidence-based test development in their test construction process.

Survey items that measure the personal beliefs, attitudes, and values of effective use of evidence-based test development processes by faculty were added by this author, aligning with the focus of this research study. These 24 items were based on review of the literature related to the current issue of evidence-based test development and its effect on nursing students' progress and success.

Ten researcher-developed survey items added to this section of the survey focus on the beliefs, values, and attitudes of faculty directly related to using evidence-based test development practices. These items aimed to explore the knowledge base, confidence level, professional responsibility, and understanding of EBP in test development. The additional intention of these items was also to identify if faculty value these practices as a significant variable in effectively and fairly evaluating their students. One item added was intended to measure the perception the faculty had regarding their nursing program's attitudes and priority towards EBP in test development. Five of the additional items developed by this researcher focused on the value and attitudes towards the use of test banks in their development process. The beliefs regarding the validity and reliability to effectively evaluate students were also addressed in these items. Two additional questions evaluate the attitudes faculty had towards the use of standardized testing in their program. These standardized tests include assessments such as the ATI® and HESI® competency focus exams. Understanding the faculty beliefs related to this type of assessment was important in uncovering the beliefs and attitudes related to effectively developing tests to prepare students to successfully pass these types of assessment tests. Four items added focused on the linguistic and cultural bias identification and faculty's attitudes towards these challenges in test development. These items were developed to measure the beliefs

and attitudes faculty have towards the cultural and linguistic biases identified as a recurring problem in nursing assessments tests. The final two questions of the survey focused on the attitudes and value the faculty place on peer review and collaboration in order to ensure that their tests meet the requirements of EBP. Measuring their beliefs and attitudes related to peer review and feedback of their tests added to the understanding of the importance of developing effective and valid tests in their nursing program. The final dissertation instrument had 124 items and, based on the field testing of the 15 nursing faculty, took approximately 30 minutes to complete.

The original survey instrument by Killingsworth (2013) was developed by utilizing several previously developed and validated assessment tools. The following section will describe the original instrument and the significance of the assessments incorporated in the development of the instrument.

The first section of the original survey had 16 items addressing credentialing, experience in teaching, current nursing program employed in, and standard demographic information. The second section addressed faculty beliefs about classroom testing. These survey items were derived from a modified version of the Ethical Climate Questionnaire (ECQ) established by Victor and Cullen (1988). The internal consistency reliability of the ECQ subscales demonstrated .60 to .80 consistency in multiple versions (Killingsworth, Kimble, & Sudia, 2015). The focus of this area is based on faculty situations and classroom testing, so the original ECQ items were modified to reflect these constructs. In addition, the original four subscales of the ECQ were combined which produced three subscales for the final survey. These three subscales were *caring, personal ethics*, *and*

rules. The scoring of this component of the survey was reflected in a one to seven Likert score. The higher the score identified, the stronger the faculty beliefs.

The next section of the survey focused on the contextual factors for the faculty's decision-making. The Evaluation of Learning Advisory Council (ELAC) survey from the NLN was used in combination with the author's additional questions, yielding 15 items within this section of the survey. The original content validity of the ELAC survey was based on assessment by15 faculty members (Oermann, Saewert, Charasika, & Yarbrough, 2009). The additional items developed by the researcher were based on the literature review of factors relevant to faculty decision making in the classroom. The results were based on a three point Likert scale with the possibility of 0 to 15 score. Again, the higher the score indicated a higher number of factors that were relevant in their decision-making process.

The final section of the survey developed by Killingsworth et al. (2015) focused on the use of best practices in classroom testing which included test construction, item analysis, and test revision. The DePew (2001) survey entitled Assessment Practice Inventory (API) and the researcher-developed items were used to develop this portion of the survey. The API was designed to determine the validity and reliability of faculty-developed test items and the correlation with the pass rates of the NCLEX-RN®. The validity statistics of the API demonstrated inter-rater reliability of .86, content validity index (CVI) for clarity of .80, and CVI for relevance of .85 in addition to moderately high to high Cronbach's alphas of .77 to .89 as reported by Zhang and Burry-Stock (2003). The researcher added items developed from the review of literature to develop the final survey used. In the final survey, 26 items related to best practices were used with a

focus on frequency calculated with 1 (not at all) to 7 (all the time) Likert-type scale. The results yielded results that determined the higher the scores, the higher the frequency the participants utilized best practices in their test development processes.

Validity

The Best Practices in Test Construction survey tool was developed under the guidance of four other accepted pre-established and validated tools as well as constructed items from Killingsworth (2013). The Assessment Practices Inventory (API), the ELAC, and the ECQ were utilized to develop this instrument. The Tarrant and Ware recommended guidelines in test construction (2011) and the review of the literature on test development were the foundation of the additional items constructed by Killingsworth (2013).

Effectively measuring the identified constructs of the survey items is of utmost importance because of the subjective nature of the study. Ensuring content validity of the Best Practices in Test Construction survey was completed by extensive review of the literature and pilot testing in order to evaluate the psychometric qualities. The author piloted the study with 34 nursing faculty from six bachelor of science nursing programs throughout the United States. The 34 pilot study participants were primarily white and female faculty members with a mean age of 49 who provided significant feedback. Based on the feedback, the instrument was revised and finalized (Killingsworth, 2013). The assessment of internal consistency was completed on the scales for seven major components of the survey ultimately demonstrating adequate reliability of the instrument.

Reliability

Reduction of measurement error is a significant component in estimating the reliability of an instrument (Kimberlin & Winterstein, 2008). All of the pre-established survey instruments used to develop the Best Practices in Test Construction survey demonstrated acceptable reliability. The following ranges were discussed: (1) Ethical Climate Questionnaire (ECQ) scales had a Cronbach's alpha range from .66 - .80 and (2) API survey had a Chronbach's alpha of .77 to .89. The dichotomization method was used to determine acceptability of the final survey (.73 - .77). The Likert-type scaling was determined to be necessary because of the potential for a larger sample in acquiring acceptable reliability (Killingsworth, 2013).

Scoring

The original survey instrument contained four sections of items with respondents rating the frequency of implementation of practices (see Appendix D). The first section focusing on the faculty beliefs about classroom testing had three subscales, *caring, personal ethics*, and *rules*, which had scoring from 1 (least descriptive) to 7 (most descriptive). Based on the scoring method, the higher the survey scores, the stronger the faculty beliefs. The second section of the survey scoring focused on the contextual factors for decision making. These 15 items were scored on three-point Likert-type scale with responses (very important, moderately important, and minimally important) based on the ELAC and researcher-developed questions. Higher scores indicated a higher number of factors that were pertinent when deciding about testing. The third section was scored based on 22 items regarding decision-making processes with a three-point Likert-type scale (not true for me, sometimes true, and true for me). The last section was scored

based on the use of best practices in the test development process. The scoring was based on 26 identified best practices. The participants were scored based on the frequency of using these best practices in their courses. The scoring frequency was on a Likert-type scale of 1 (not at all) to 7 (all the time). The higher the score, the greater the use of the particular practices.

An additional 24 researcher-developed questions were added to the original survey by this researcher. These items specifically focused on the beliefs, attitudes, and values with the respondents indicating 1 (strongly disagree) to 7 (strongly agree) for each particular practice element. The higher the score, the more likely the participant had positive beliefs, attitudes, and values regarding evidence-based test development.

General Statistical Strategy

Descriptive statistics utilize data collection and analysis techniques that yield reports concerning the measures of central tendency, variation, and correlation. The combination of its characteristic summary and correlational statistics, along with its focus on specific types of research questions, methods, and outcomes is what distinguishes descriptive research from other research types (Punch, 2005).

The statistical strategy of this study began with a descriptive analysis to demonstrate the distribution of responses across the sample and for sub-groups within the sample. The study also aimed to identify measures of central tendency including mean, median, mode, deviance from mean, variation, and percentage. The data were analyzed using frequency distributions and cross tabulations and displayed using tables and charts. All of the analyses were conducted using the Statistical Package for the Social Sciences® (SPSS).

The statistical data analysis of this study was aimed to answer the three research questions previously discussed. The first research question aimed to identify the features of test construction that were utilized by faculty in pre-licensure programs. Data from these items of the survey instrument focused on *test construction* and the descriptive statistics derived from the data collected. The second research question's objective was to identify the features of the test revision process that were utilized by nursing faculty in pre-licensure programs. Data from these items of the survey instrument focused on *test revision*, and the additional descriptive statistics derived from the data collected intended to answer this question. The differences-between ASN and BSN faculty in using best practices in test development were analyzed using all items.

Data Cleaning

The data cleaning process is an essential aspect of quality assurance and a determinant of study validity. The data cleaning process for this study followed the data cleaning framework presented by Van den Broeck, Cunningham, Eeckels, and Herbst (2005). This process has three phases involving repeated cycles of screening, diagnosing, and editing of suspected abnormalities.

The first phase of the process is screening. Using SPSS, the researcher identified suspected data and flagged dubious data, patterns, or results. The researcher, within this phase, checked the surveys and validated the data retrieved from the Research Electronic Data Capture (REDCap) system. Frequency distributions and cross-tabulations were used and statistical outliers identified. The second phase of the data cleaning is the diagnostic phase where the researcher clarified the true nature of the questionable data points, patterns, and statistics. The final phase of the process is the treatment phase where the

researcher decided what to do with the problematic observations. The questionable data were then either corrected, deleted, or left unchanged, depending on the impact and influence on the study.

Descriptives

Descriptive statistics help to understand the experiment or data set in detail and demonstrate the required details that help put the data in perspective (Polit & Beck, 2008). The use of descriptive statistics in this study utilized numerical and graphical methods to identify patterns in the data set. It was also used to summarize information discovered in the data set and present the findings in a convenient form (Polit & Beck, 2008). This will be provided in chapter 4.

Reliability Testing

The accuracy and consistency of data collected in a study ensure its reliability (Polit & Beck, 2008). The data in this study were collected utilizing the REDCap® online commercial platform which helped in ensuring the reliability of the results. Cronbach's alphas were calculated for the competencies used within this study (e.g., internal consistency).

<center>**Limitations**</center>

Threats to Internal Validity

The internal validity is used to measure the accuracy and the soundness of the study. It is the extent to which a research study is free from errors and any difference in measurement is due to the independent variable and nothing else. Internal validity can be affected by many factors (Polit & Beck, 2008). History can affect internal validity by having an unanticipated event occur during the study to affect the dependent variable. At

the point of this study completion, no major events had been identified as potentially affecting the internal validity of this study. Maturation, which focuses on the changes in the dependent variable due to function of time, is another threat to internal validity. However, in this study, the maturation is not considered a threat since the survey was not expected to take longer than 30 minutes to complete and only required a one-time response. This research study did not utilize a pre-test design, and the data were gathered only once from each of the participants. Therefore, the testing and instrumentation threats to internal validity were considered minimal. Selection error is another threat to internal validity. The researcher controlled this threat by asking the participants to read the informed consent and only participate in the study if they had met the inclusion criteria and did not meet the exclusion criteria. It was hopeful that those who participated in the study followed these directions and participated accordingly. Finally, the statistical conclusion validity was ensured by the use of the adequate power. In this research study, the power was set at 0.80. Utilizing the proper power ensured that relationships were detected among variables within this study.

Threats to External Validity

The external validity is used to identify the correctness of the research findings, by examining their applicability from one setting to another. It is the extent to which the research results can be inferred for the world at large. It determines whether the causal relationship discovered in the study can be generalized or not (Polit & Beck, 2008). Generalization of this research study extended to academic nurse educators teaching in pre-licensure undergraduate nursing programs, regardless of their educational status, certification status, or years of faculty experience. In comparison to the national

candidate performance on the NCLEX-RN®, for 2017, Florida had a 75.1% pass rate for their first-time test takers of states with similar number of candidates such as Texas, New York, Pennsylvania, and Ohio. A threat to external validity was the existence of various educational programs nationally and the demographic characteristics of this study population which may have limited generalizability.

Chapter Summary

This chapter detailed the methodology and statistical analysis that was used in this descriptive, non-experimental study of how the beliefs, attitudes, and values of nursing faculty affect their use of evidence-based practices in test development. The setting and sampling plan focused on obtaining a non-probability, purposive sampling format by using the deans and directors of nursing programs as gatekeepers to disseminate the survey to the faculty. The eligibility criteria focused on faculty who currently create tests for their nursing classes. The sample size was determined by the Modified Cochran formula, and a total of 117 surveys were returned. The self-reporting survey provided descriptive data regarding test construction practices including decision-making processes, test development, test revision, test analysis, and beliefs, attitudes, and values regarding EBP in test development. The validity and reliability of the instrument were extensively discussed. The data collection method by utilizing REDCap® survey and data collection platform and analysis plan using SPSS® were also addressed. This study's limitations relating to its internal and external threats were addressed to identify components that may be a threat to the overall validity of the study. Overall, the researcher uncovered contributing factors and variables that impact evidence-based testing practices in nursing education.

Chapter Four

Results

Utilizing best practices in test development is a significant element in nursing education. Invalid assessment processes produce invalid data regarding student competency. Faculty test development practices should adhere to best practices in order to effectively prepare students for safe clinical practice. The purpose of this quantitative study was to identify and describe faculty beliefs, attitudes, and values regarding evidence-based test development in pre-licensure nursing programs in the state of Florida. Data collection was opened on November 5, 2018, and closed on December 15, 2018. Analysis of the data was completed using SPSS version 25 and will be discussed further in this chapter. This chapter will also include the data collection processes of the study, the description of the sample and response measurements, and the results of the data analysis.

Data Cleaning

Data were obtained by an Internet-based survey sent to pre-licensure nursing faculty in the state of Florida. The survey was delivered and accessed by the survey platform REDCap. Raw data were collected and imported from REDCap to SPSS version 25 to create the database. Cleaning of the data included identifying and eliminating incorrect and/or inaccurate information and rejecting incomplete surveys. All 117 surveys were analyzed, and two were discarded because of incomplete information. Eleven (11)

survey participants did not meet the requirements, so the final number for the sample size was 104. No additional outliers were found.

To answer Research Question 1: What features of test construction are utilized by nursing faculty in pre-licensure programs? descriptive analysis was used to analyze the data from the *Test Construction* and the *Beliefs, Attitudes, and Values* sections of the survey. To answer Research Question 2: What features of test revision are utilized by nursing faculty in pre-licensure programs? the *Test Revision* section of the survey data were analyzed and presented. To answer Research Question 3: What are the differences in test development practices in pre-licensure programs between ASN and BSN faculty? an independent sample t-test was used to analyze the data.

Descriptives

Description of the Sample

The participants in this study were pre-licensure nursing faculty who participated in test development and/or item writing within their nursing programs. A total of 95 nursing programs were emailed, and the final number of respondents was 117. Of the 117 faculty who participated, two surveys were incomplete and 11 faculty did not continue past the screening questions. The final sample size for this study was 104.

Demographic information was obtained from the participants to identify specific variables related to the sample population. The data obtained depicted the predominant nursing program type being BSN ($n = 46$, ASN; $n = 58$, BSN). The area code 305 had the most responses which was from the city of Miami ($n = 19$). The most frequent number of years the nursing faculty had been writing test items was 4-7 years ($n = 38$). The method of test delivery that was most prevalent was a combination of computer and paper/pencil

testing ($n = 8$, paper and pencil; $n = 35$, computer; $n = 61$, both). The MSN was the most

frequently reported degree held by the participants (n = 85, MSN; $n = 3$, BSN; $n = 16$,

doctorate). MSN with no specialty area was the most reported specialization by the

participants ($n = 39$). Most of the participants reported having received no courses in test

development within their nursing employment institutions ($n = 78$). The majority of

participants responded negatively to having participated in a professional development

program focusing on test development ($n = 59$). Most respondents reported that they did

not hold certification as a nurse educator ($n = 82$). The participants were asked if there

had been any concerns expressed by their administration about the NCLEX-RN® pass

rates for their programs. The majority of the respondents affirmed that there had been

discussion of issues with pass rates within their nursing program ($n = 93$).

Table 3

Frequencies and Percentages of Sample Demographic Data

Variable	Category	N	%
Program type	ASN	46	42.3
	BSN	58	56.7
	Both	1	1
Area code	305	19	18.3
	954	15	14.4
	941	14	13.5
	407	12	11.5
	904	11	10.6
	561	8	7.7
	813	4	3.8

	239	3	2.9
	Others	18	45.2
Years writing tests	1-3	23	22.1
	4-7	38	36.5
	8-10	26	25
	11-15	12	11.5
	16 years or more	5	4.8
Test delivery method	Paper/pencil	8	7.7
	Computer	35	33.7
	Both	61	58.7
Highest degree	BSN	3	2.9
	MSN	85	81.7
	Doctorate	16	15.4
Specialization	Nurse Educator	29	27.9
	Nurse Administrator	11	10.6
	Nurse Practitioner	19	18.3
	Clinical Nurse		
	Specialist	6	5.8
	None of the above	38	36.5
	Do not have master's	1	1.0
Courses in test	No course	78	75
development	One course	13	12.5
	More than one course	13	12.5
Professional	Yes	44	42.3
development in test	No	59	56.7

development			
participation			
	Cannot remember	1	1
Certification as a nurse	No	82	78.8
educator			
	Yes	22	21.2
Concerns about	Yes	93	88.4
NCLEX pass rates	No	10	9.6
	Do not know	1	1.0

Note. N = 104.

Reliability Testing

Identifying reliability of an instrument is paramount in a research study. The reliability of the data collected and the methodology used must be high in order to provide sufficient evidence upon which decisions are based (Downing, 2004). The instrument used in this study was originally based on the Best Practices in Test Development created by Killingsworth (2013). Additional questions were added by the primary investigator of this study, and this tool was used to answer all three research questions.

Table 4 illustrates the Cronbach's alpha for the three scales used in this study. The *Faculty's Beliefs, Attitudes, and Values of Evidence-Based Test* scale is satisfactory in terms of internal coherency (criteria of Cronbach's Alpha ≥ 0.70). The *Test Revision* and *Components in Test Construction* scales have lower consistency than the ideal for research purposes. This issue could be explained because the construct intended to evaluate is heterogeneous (it is evaluating different aspects of test revision and

components of test construction), causing the internal consistency to be slightly lower than the ideal. However, these scores are in range with Killingsworth's (2013) original scales.

Table 4

Scales Reliability

Items	Cronbach's Alpha	N of Items
Faculty's Beliefs, Attitudes and Values of Evidence-Based Test	0.78	24
Test Revision	0.61	11
Components Test Construction	0.64	13

Research Questions

This study contained three research questions. The data collected from the Internet-based survey provided the results to all research questions. A total of 117 faculty members participated in the survey with 104 being the final sample size. The demographic data collected from the survey provided information about the sample. The frequencies and percentages as well as the descriptive statistics of the respondent data are described in the next sections.

Research Question 1

The first research question was the following: What features of test construction are utilized by nursing faculty in pre-licensure programs? Item level descriptive statistics were performed on the *Test Construction* and the *Beliefs, Attitudes, and Values* sections of the survey data. For each of the questions in the tables below, the participants answered according to a Likert-type scale (1-7) ranging from l*east descriptive to most*

descriptive in Table 5 and s*trongly disagree to strongly agree* in Table 6. The means in

the Table 6 indicate how frequently the participants used the best practice described in

their test development processes. The means in Table 6 indicated the level of agreement

with the beliefs described.

Table 5

Descriptive Statistics for Faculty Use of Best Practices in Test Construction (ranked by the mean)

n°	Item	Mean (SD)		Median
3	Major content topics	5.48	(1.52)	6
8	Higher cognitive levels according to Bloom's taxonomy (e.g., application, analysis, evaluation)	5.02	(1.92)	6
13	Overall, how skilled are you at test construction?	4.96	(1.95)	6
12	Use various test item types (e.g., multiple-choice, select all that apply, fill in the blank)	4.76	(2.09)	6
4	Specific content topics	4.57	(2.19)	6
2	Class or unit objectives	4.38	(1.99)	5
9	Clinical context for test items.	4.32	(2.24)	4
10	Plausible distractors in multiple-choice test items	4.22	(2.30)	6
11	Even distribution of correct answer in multiple-choice options	4.16	(2.28)	6
1	Course objectives	4.14	(2.00)	4
5	A test blueprint or table of specifications	4.02	(1.95)	4
6	The NCLEX-RN® test plan	3.67	(1.98)	4
7	Peer review of test items	3.08	(2.00)	3

Table 6

Descriptive Statistics for Faculty Beliefs, Attitudes, and Values of Evidence-Based Test

Development (ranked by the mean)

nº	Item	Mean	(SD)	Median
20	Linguistic and cultural bias is a common problem when using test banks.	5.37	(1.87)	6
22	I value feedback from my colleagues on tests I have developed.	5.33	(1.99)	6
4	Knowledge of evidence-based test development is important to me.	5.22	(2.33)	7
1	Evidence-based test development is an essential part of my role as a nurse educator.	5.15	(1.88)	6
16	Evidence-based test development skills are useful but not a necessary component of pre-licensure nursing education.	4.81	(1.60)	5
21	It is important to me to have other faculty review my test prior to administration.	4.74	(2.08)	5
8	Using test bank items saves me time in the test development process.	4.70	(2.12)	6
2	I understand the evidence-based practices in test development.	4.62	(1.94)	5
14	Faculty-made tests are not as important as standardized tests (e.g., HESI, ATI) in preparing for the NCLEX-RN®.	4.42	(1.81)	5
6	It is the responsibility of all faculty members to participate in evidence-based test development.	4.34	(2.46)	2
11	Test development using evidence-based practices effectively prepares students for NCLEX-RN®.	4.26	(2.33)	6

10	Standardized tests identify the students that should/should not progress in the nursing program.	4.23	(1.51)	4
5	I am confident in my ability to prepare evidence-based test items.	4.20	(1.93)	4
3	I always implement evidence-based practices in my test development process.	4.10	(1.67)	4
7	The more time I spend developing evidence-based test items, the less time is available for other critical educational duties.	4.08	(2.25)	5
12	Test development using evidence-based practices effectively prepares students for standardized tests (e.g., HESI, ATI).	4.08	(2.41)	2
9	Test bank items are just as effective as faculty-developed items.	3.91	(1.91)	4
18	I often restructure test bank items to change their difficulty.	3.78	(2.33)	5
15	I place more emphasis on evidence-based teaching than on evidence-based test development.	3.77	(2.30)	4
13	Evidence-based test development is a skill that all nursing faculty must have.	3.34	(2.46)	2
23	My nursing program places a priority on evidence-based test development.	3.16	(1.69)	3
19	Always evaluating my tests for linguistic and cultural bias is important in my test development process.	2.80	(2.27)	1
17	Most test banks are valid and reliable.	2.44	(1.66)	3

Research Question 2

The second question asked the following: What features of test revision are utilized by nursing faculty in pre-licensure programs? To answer this question, the item

level descriptive statistics were reported. The faculty answered the items on a Likert-type scale (1-7) from *not at all* to *all the time*. The means in Table 7 indicated the frequency the faculty performed these actions during tests revision.

It can be seen that the faculty most reported changing test items to ensure test security (M = 5.24), using item analysis data when determining to keep or eliminate test questions before finalizing test scores (M = 5.03), assessing for outdated language used in test items (M = 4.87), and comparing item analysis data for test questions used repeatedly from one term to another (M = 4.82).

Table 7

Descriptive Statistics for Faculty Use of Best Practices in Test Revision (ranked by the mean)

n°	Item	Mean	(SD)	Median
8	Change test items to ensure test security.	5.24	(1.66)	6
11	Overall, how skilled are you at test revision?	5.11	(1.86)	6
1	Use item analysis data when determining to keep or eliminate test questions before finalizing test scores.	5.03	(1.96)	6
7	Assess for outdated language used in test items.	4.87	(1.87)	5
2	Compare item analysis data for test questions used repeatedly from one term to another.	4.82	(1.87)	5
10	Change test items to ensure sufficient sampling of content.	4.78	(2.09)	6
4	Use difficulty level of test items to revise test items.	4.66	(2.03)	5
9	Change test items to reflect emphasis in classroom content.	4.57	(2.10)	6
3	Use distractor discrimination to revise test items.	3.39	(2.02)	3
6	Assess for changes in domain content based upon new research data.	3.08	(1.92)	3

| 5 | Assess for linguistic/cultural bias in test items. | 2.67 | (2.05) | 2 |

Research Question 3

The third question asked the following: What are the differences in test development practices in pre-licensure programs between ASN and BSN faculty? To answer this question, the means and standard deviation were identified and then an independent t-test was used. There were 46 ASN faculty respondents, 58 BSN faculty respondents, and one respondent who taught in both types of programs. The differences between the ASN and the BSN faculty practices are depicted in Table 8.

The independent t-test was used to analyze the differences between the scores of all three subscales between ASN and BSN faculty in Tables 9 through 11. The total scores of the subscales did not show significant differences in the faculty responses; however, some items individually showed significant differences in responses between the two groups.

There were significant differences in items 5, 6, and 13 on the *Test Construction subscale* which is indicated by a *p ≤ 0.05*. Item 5 demonstrated ASN (M = 4.30, SD = 1.79) and the BSN (M = 5.10, SD = 2.0); *t = 100, p = 0.03*. Item 6 demonstrated ASN (M = 3.19, SD = 1.85) and BSN (M = 3.98, SD = 1.99); *t = 100, p = 0.04*. Item 13 demonstrated ASN (M = 4.30, SD = 1.94) and BSN (M = 5.10, SD = 1.90); *t = 100, p = 0.02*. The *Test Revision subscale* indicated two items had significant differences in responses. Item 3 demonstrated the ASN (M = 3.27, SD = 2.00) and the BSN (M = 4.24, SD = 2.03); *t = 100, p = 0.04*, and item 6 demonstrated the ASN (M = 3.16, SD = 1.76) and the BSN (M = 4.22, SD = 2.05); *t = 100, p = 0.03*. Finally, the *Faculty Beliefs, Attitudes, and Values subscale* items showing significant differences were items 6, 8, 13,

and 19. Item 6 demonstrated the ASN (M = 4.16, SD = 2.47) and the BSN (M = 5.36, SD

= 2.49); *t=100, p=0.03*. Item 8 demonstrated the ASN (M = 4.20, SD = 2.06) and the

BSN (M = 5.30, SD = 2.20); *t = 100, p = 0.02*. Item 13 demonstrated the ASN (M =

3.74, SD = 2.47) and the BSN (M = 4.92, SD = 2.42); t = 100, p = 0.04. Item 19

demonstrated the ASN (M = 2.30, SD = 1.96) and the BSN (M = 3.20, SD = 2.47), *t =*

100, p = 0.05.

Table 8

Means for Faculty Beliefs, Attitudes, and Values of Evidence-Based Test Development of ASN

and BSN Faculty

n°	Item	ASN	BSN
1	Evidence-based test development is an essential part of my role as a nurse educator.	5.19	5.11
2	I understand the evidence-based practices in test development.	4.70	4.61
3	I always implement evidence-based practices in my test development process.	4.07	4.07
4	Knowledge of evidence-based test development is important to me.	5.00	5.45
5	I am confident in my ability to prepare evidence-based test items.	4.45	3.96
6	It is the responsibility of all faculty members to participate in evidence-based test development.	4.26	4.45
7	The more time I spend developing evidence-based test items, the less time is available for other critical educational duties.	4.26	4.00
8	Using test bank items saves me time in the test development process.	4.77	4.63
9	Test bank items are just as effective as faculty-developed items.	3.70	4.02

10	Standardized tests identify the students that should/should not progress in the nursing program.	4.36	4.11
11	Test development using evidence-based practices effectively prepares students for NCLEX-RN®.	4.21	4.36
12	Test development using evidence-based practices effectively prepares students for standardized tests (e.g., HESI, ATI).	4.19	3.95
13	Evidence-based test development is a skill that all nursing faculty must have.	3.74	2.98
14	Faculty-made tests are not as important as standardized tests (e.g., HESI, ATI) in preparing for the NCLEX-RN®.	4.57	4.37
15	I place more emphasis on evidence-based teaching than on evidence-based test development.	3.60	3.95
16	Evidence-based test development skills are useful but not a necessary component of pre-licensure nursing education.	4.65	4.95
17	Most test banks are valid and reliable.	2.60	2.27
18	I often restructure test bank items to change their difficulty.	4.05	3.54
19	Always evaluating my tests for linguistic and cultural bias is important in my test development process.	2.30	3.20
20	Linguistic and cultural bias is a common problem when using test banks.	5.45	5.30
21	It is important to me to have other faculty review my tests prior to administration.	4.49	4.95
22	I value feedback from my colleagues on tests I have developed.	5.23	5.41
23	My nursing program places a priority on evidence-based test development.	2.86	3.32

Table 9

T test (Independent Samples) for Test Construction Subscale

		ASN			BSN			t-test		
Items	**n**	**M**	**SD**	**n**	**M**	**SD**	**M. dif**	**t**	**df**	**p**
1	46	4.09	1.89	58	4.21	2.10	-0.12	-0.29	100	0.77
2	46	4.34	1.83	58	4.45	2.12	-0.11	-0.27	100	0.79
3	46	5.80	0.95	58	5.32	1.73	0.47	1.77	100	0.08
4	46	4.50	2.24	58	4.60	2.19	-0.10	-0.22	100	0.83
5	46	4.30	1.79	58	5.10	2.00	-0.80	-0.40	100	*0.03
6	46	3.19	1.85	58	3.98	1.99	-0.80	-2.04	100	*0.04
7	46	2.84	1.84	58	3.21	2.11	-0.37	-0.92	100	0.36
8	46	5.05	1.88	58	4.98	2.00	0.06	0.16	100	0.87
9	46	4.40	2.15	58	4.25	2.34	0.15	0.33	100	0.74
10	46	4.47	2.20	58	4.00	2.38	0.47	1.01	100	0.31
11	46	4.56	2.12	58	3.83	2.38	0.73	1.63	100	0.11
12	46	5.00	1.90	58	4.65	2.20	0.35	0.84	100	0.41
13	46	4.30	1.94	58	5.10	1.90	-0.80	-0.80	100	*0.02
Total										
Score	**46**	**57.23**	**9.70**	**59**	**56.25**	**11.10**	**0.97**	**0.46**	**100**	**0.64**

Note. M dif = Mean of differences among (ASN) and (BSN) *$p \leq .05$.

Table 10

T test (Independent Sample) for Faculty Beliefs, Attitudes, and Values of Evidence-Based Test

Development Subscale

	ASN			BSN			t-test			
Items	**n**	**M**	**SD**	**n**	**M**	**SD**	**M. Dif**	**t**	**df**	**p**
1	46	5.19	1.87	58	5.11	1.93	0.08	0.20	100	0.84
2	46	4.40	1.98	58	5.37	1.92	-0.09	0.23	100	*0.02
3	46	4.07	1.64	58	4.07	1.68	0.00	0.00	100	1.00
4	46	5.00	2.35	58	5.45	2.30	-0.45	-0.95	100	0.35
5	46	4.45	1.97	58	3.96	1.88	0.49	1.25	100	0.21
6	46	4.16	2.47	58	5.36	2.49	-1.20	-0.42	100	*0.03
7	46	4.26	2.17	58	4.00	2.31	0.26	0.56	100	0.58
8	46	4.20	2.06	58	5.30	2.20	-1.10	0.39	100	*0.02
9	46	3.70	1.71	58	4.02	2.04	-0.32	-0.83	100	0.41
10	46	4.36	1.48	58	4.11	1.55	0.25	0.81	100	0.42
11	46	4.21	2.38	58	4.36	2.30	-0.15	-0.33	100	0.75
12	46	4.19	2.45	58	3.95	2.40	0.24	0.49	100	0.62
13	46	3.74	2.47	58	4.92	2.42	-1.18	1.54	100	*0.04
14	46	4.57	1.93	58	4.37	1.70	0.20	0.56	100	0.58
15	46	3.60	2.20	58	3.95	2.37	-0.35	-0.75	100	0.45
16	46	4.65	1.59	58	4.95	1.63	-0.30	-0.91	100	0.37
17	46	2.60	1.92	58	2.27	1.42	0.33	0.98	100	0.33
18	46	4.05	2.32	58	3.54	2.35	0.50	1.07	100	0.29
19	46	2.30	1.96	58	3.20	2.47	-0.90	-2.01	100	*0.05
20	46	5.45	1.78	58	5.30	1.97	0.15	0.40	100	0.69

Items	n	M	SD	n	M	SD	M Dif	t	df	p
21	46	4.49	2.15	58	4.95	2.04	-0.46	-1.09	100	0.28
22	46	5.23	2.07	58	5.41	1.97	-0.18	-0.44	100	0.66
23	46	2.86	1.55	58	3.32	1.72	-0.46	-1.36	100	0.18
Total Score	**46**	**95.63**	**11.50**	**58**	**95.88**	**9.60**	**-0.25**	**-0.12**	**100**	**0.91**

Note. M dif =Mean of differences among (ASN) and (BSN) *$p \leq .05$.

Table 11

T-test (Independent Sample) for Faculty Use of Best Practices in Test Revision Subscale

Items	ASN			BSN			t-test			
	n	M	SD	n	M	SD	M Dif	t	df	p
1	46	5.09	1.92	58	4.97	2.01	0.13	0.32	100	0.75
2	46	4.49	1.92	58	5.12	1.78	-0.63	-1.70	100	0.09
3	46	3.27	2.00	58	4.24	2.03	-0.07	-0.72	100	*0.04
4	46	4.58	2.07	58	4.67	2.01	-0.09	-0.22	100	0.82
5	46	2.37	1.93	58	2.81	2.06	-0.44	-1.08	100	0.28
6	46	3.16	1.76	58	4.22	2.05	-0.06	-0.46	100	*0.03
7	46	4.79	1.87	58	4.90	1.88	-0.11	-0.28	100	0.78
8	46	5.23	1.64	58	5.24	1.70	-0.01	-0.04	100	0.97
9	46	4.93	1.84	58	4.28	2.25	0.65	1.59	100	0.12
10	46	4.68	2.11	58	4.91	2.06	-0.23	-0.55	100	0.58
11	46	5.14	1.86	58	5.05	1.88	0.08	0.23	100	0.82
Total Score	**46**	**47.07**	**5.58**	**58**	**48.22**	**4.83**	**-1.16**	**-1.12**	**100**	**0.27**

Note. M dif = Mean of differences among (ASN) and (BSN) *$p \leq .05$.

Chapter Summary

This chapter documents the research findings for each of the three research questions in this study. The data collected were analyzed using SPSS version 25. To answer research questions one and two, descriptives, frequencies, and percentages were presented. A regression correlational analysis was used to analyze the data to address research question three. The data collected demonstrated the current practices and beliefs of nursing faculty during test development processes. The summary of the findings, implications, and recommendations will be discussed in Chapter 5.

Chapter Five

Discussion and Summary

The aim of this study was to gain insight into the test development practices of nursing faculty in pre-licensure programs in the state of Florida and to add to the knowledge about beliefs, attitudes, and values about evidence-based test development practices in nursing education. Results of this study may assist nurse educators and nursing programs to improve test development processes which impact student success and improve preparation for professional practice. Rokeach's beliefs, attitudes, and values theory provided the framework and the post-positivist worldview guided the research. This chapter will summarize the research findings, discuss the implications of this quantitative study, integrate the literature previously presented with the results of this study, and discuss the limitations and recommendations for future research.

Summary of Findings

Research Question 1

Research question 1 asked: What features of test construction are utilized by nursing faculty in pre-licensure programs? This question was answered by the *Test Construction* and the *Beliefs, Attitudes, and Values in Test Development* sections of the survey. The results demonstrated that faculty are not consistent with best practices in test development in their nursing programs. Their beliefs and attitudes towards evidence-based test development indicate a concern with their understanding and implementation of evidence-based practices. Most faculty revealed that areas such as linguistic and

cultural bias were not addressed within their test development process. It is important to highlight that utilizing test banks was a significant practice demonstrated in test development by the faculty even though the majority of them strongly disagreed that the test banks were valid and reliable.

Another component that showed notable performance was the time factor in test development practices. Faculty reported that developing evidence-based test items was time consuming and that they perceived it led to neglecting other critical educational duties. Most faculty did not engage in peer review of test items, utilize the NCLEX-RN® test plan, or utilize test blueprinting. However, most faculty did indicate that they utilized course objectives, class/unit objectives, and major content areas to develop their course tests. The higher cognitive levels of Bloom's taxonomy and various test item types were notably used frequently by the majority of faculty.

Most faculty strongly agreed that evidence-based test development is an essential part of their role as a nurse educator; however, they strongly disagreed to using evidence-based practices in developing their course tests regularly. Most faculty strongly agreed with the importance of standardized tests; however, the majority of faculty denied using EBP to prepare the students for these standardized tests. Most faculty strongly disagreed that their nursing programs place a priority on EBP in test development. Overall, the majority of faculty felt they were not very skilled at evidence-based test construction practices.

Integration of the Findings with Previous Literature

Reasons for the lack of integration of evidence-based practice in nursing education appear to be related to the lack of understanding and training in evidence-based

test construction practices and proper item development. This finding is in line with Clifton and Schriner (2010) who identified that tests developed by faculty are relatively low quality in item structure and objectives and can significantly improve by providing evidence-based test development training. The availability of a sufficient number of faculty development opportunities focused on assessment are essential for recruiting and retaining faculty commitment (Abate, Stamatakis, & Haggett, 2003). Faculty workshops and routine inservice training are the most common professional development methods.

The majority of the faculty in this study confirmed they did not use the blueprinting and utilization of course and class objectives as frequently in their test development process. This finding was consistent with the data reported by Killingsworth (2013). According to Hamdy (2007), an important aim of blueprinting is to reduce two significant threats to validity. The first threat is under-representation and under-sampling of the course content and objectives. The second threat is construct-irrelevant variance. This may be present in tests with errors in item formats, items that are flawed in difficulty levels, and items with linguistic and cultural bias. Although test blueprinting is an efficient method for enhancing the test development process, its application has been discounted by nurse educators (Hicks, 2011).

The amount of time needed to develop valid items was an issue for the faculty in this study. Most faculty felt that the time spent on evidence-based test development took away time from other important educational interventions. This is in line with the findings by Hicks (2011) where faculty issues with developing well-constructed test items were directly related to the time constraints of the process. Lack of preparation and

lack of sufficient time for effective test construction and analysis were deterrents to best practices.

Evaluating tests for linguistic and cultural bias was an area that was not performed by the majority of the faculty in this study. They also showed that they did not assess the test items used from the test banks for these biases either. This is an essential component of ensuring a validity of an item or a test. It is possible that faculty do not feel competent in identifying cultural and linguistic bias in test items, due to lack of training, so this practice is often ignored. These findings were in line with Killingsworth's (2013) study. Klisch (1994) reinforces this practice by describing the effects that any type of bias can have on the validity of the test results. These biases can lead to confusing language and construct-irrelevant variances which can affect the student's performance on the item. Identifying biases in test items creates a fair test in which students of equal ability are likely to answer the item correctly, leading to valid data and evaluation of student competence.

Research Question 2

Research question 2 asked: What features of test revision are utilized by nursing faculty in pre-licensure programs? This question was answered by the *Test Revision* section of the survey. The results demonstrated that the majority of faculty consistently used item analysis to determine if an item was to be eliminated or retained before finalizing the test scores. The faculty did not compare the item analysis data for questions from one term to another, nor did they use distractor discrimination to revise test items.

The majority of faculty respondents do not routinely assess for linguistic/cultural bias, assess for changes in domain content based upon current research data, or assess for

outdated language in test items in the revision process. Test security was an area that most faculty had significant focus. The majority answered that they changed test items to ensure test security *all the time*. They also answered positively to changing items to reflect emphasis in classroom content and ensuring sampling of content. Overall, faculty indicated that they felt very skilled in the test revision process.

Integration of the Findings with Previous Literature

Although faculty indicated they were skilled at the test revision process, there were key steps in the process to which they responded negatively. The faculty focused on certain components of the revision process while overlooking other significant procedures. It is possible that the missing steps in the test revision process are related to the time constraints as well as the lack of understanding of evidence-based test development practices as discussed previously. Research has described the importance of devoting sufficient time to test revision practices prior to and after administration of a test (Haladyna & Downing, 1985; Haladyna & Downing, 1989a; Morrison, Nibert, & Flick, 2006; Tarrant et al., 2006; Tarrant & Ware, 2011). In addition, time pressures force faculty to create tests hurriedly, sometimes even hours before a test is administered. Pre- and post-administration review and revision are often neglected, and a poor quality test is the end product (Clifton & Schriner, 2010; Tarrant et al., 2006).

The revision process following the administration of the test is important as it focuses on the item performance analysis statistics. The majority of faculty responses suggested that the faculty focused on test security of the exam as a priority in the revision process. This focus on test security by faculty respondents may be related to the increased concern with academic integrity in nursing programs. According to Cascoe, Stanley,

Stennett, and Allen (2017), there has been a prevalent challenge in nursing programs with academic dishonesty. Cheating behaviors in nursing programs are common and have been supported by literature. These behaviors include cheating on an examination, plagiarism, or forgery and can be in any setting in the nursing program such as clinical, lab, or didactic (Kreueger, 2014; McCabe, 2009, Theart & Smit, 2012).

It was also identified that nursing faculty considered analysis of correct items and distractors chosen to be an important step in the revision process. The majority of the faculty who participated in the study also responded that they do not analyze the performance of the higher performing students compared to the lower performing students when reviewing and revising a test. This component of the test analysis process is essential to establishing validity of test items and the ability to discriminate between the types of students and their performance (Zaidi et al., 2018).

Completing a distractor analysis is also an area that showed inadequate performance. The distractor analysis is a significant element in determining the quality of the test item and, specifically, the quality of the items. If faculty were just to assess the item difficulty and the item discrimination index, this would not be effective in assessing the performance of the distractors. So this component of the test revision process is essential in addressing the plausibility of the item and the ability to discriminate between the students' abilities (Hicks, 2011). Based on the responses from the faculty participants, it is safe to conclude that a thorough revision process is not incorporated into their test development practices.

Research Question 3

Research question 3 asked: What are the differences in test development practices in pre-licensure programs between ASN and BSN faculty? This question was answered from the *Test Construction* and the *Beliefs, Attitudes, and Values in Test Development* sections of the survey. Using the independent sample t-test, there was minimal difference between ASN and BSN faculty regarding use of evidence-based practices in test construction within their nursing programs. As discussed previously, the differences among means between groups are small, which supports the assumption that ASN and BSN faculty have similar test development practices. There were certain items in each subscale, however, that demonstrated significance in different responses of the participants. ASN faculty did not feel as strongly about evidence-based practice in test development being a necessary skill for nursing faculty in pre-licensure programs. BSN faculty believed that it was the responsibility of the nursing faculty to participate in evidence-based test development more than ASN faculty did. Both ASN and BSN faculty responded to using test banks in their test development process; however, ASN faculty felt more strongly that using test banks saved them time in their test development process. Evaluating linguistic and cultural bias in test items was determined to be a process that was not consistently used by either ASN or BSN faculty; however, ASN faculty did not value the process of assessing for these biases as much as the BSN faculty.

Based on the analysis of the *Test Construction subscale*, ASN faculty did not use the test blueprinting process or the NCLEX test plan as frequently as the BSN faculty in developing their tests. Regarding overall skill of constructing evidence-based tests, BSN

faculty felt more skilled than the ASN faculty. Although both ASN and BSN faculty responded negatively to using distractor discrimination in test revision, the BSN faculty demonstrated applying this technique significantly more than the ASN faculty. Using new research data to revise tests was a process that the BSN faculty responded more positively to than the ASN faculty, demonstrating significant differences.

Although there have been several items that determined a difference in test development processes between the ASN and BSN faculty, the overall responses were similar between the two groups of educators. The differences may be attributed to other variables which can affect these results such as specific program requirements, experience of the faculty, and personal perceptions.

Integration of the Findings with Previous Literature

The ASN degree is generally a two-year program of study but can range from two to four years after the student has completed all prerequisite general education and science courses (Aiken et al., 2011, Billings & Halstead, 2011). These programs have been integral parts of the community college setting. Within the past few years, accelerated ASN programs have developed which allow students to complete their degrees in under two years. As with the BSN program, the curriculum is designed to prepare the students to pass the NCLEX-RN® and to meet the safety requirements of professional practice (Duphily, 2011). Similar to their BSN counterparts in college or university settings, ASN faculty must understand the foundations of nursing education including curriculum models, assessment and evaluation methods, and instructional methodologies. Although the programs are two years, the minimum education required to teach in these programs is a graduate degree (Duphily, 2011).

The NCSBN (2008) published a document produced by its Faculty Qualifications Committee discussing the qualifications and roles of nurse educators in nursing education. The committee recommended that all nursing faculty in pre-licensure programs including ASN and BSN be graduate prepared in nursing which includes clinical practice, teaching and learning, and curriculum development and implementation. This is in line with the number of respondents of this dissertation study holding master's degrees as their highest degree achieved. Based on the fact that MSN faculty are the most prominent degreed faculty in both ASN and BSN pre-licensure nursing programs, it is safe to postulate that this contributes to the similarity in test development practices.

Most nurse educators enter into the academic environment after years of clinical practice and having earned an MSN. In light of that, it is likely that the transition process experienced by novice nurse educators and acclimating to their new role are challenging. During this transition, whether in an ASN or BSN program, few new nurse educators are exposed to best practices in test development. This could be a contributing factor to the similarities in responses related to test development practices. In addition, faculty, both ASN and BSN, responded to not engaging in many professional development opportunities related to test development. Resources for new faculty and senior nurse educators are often limited regarding evidence-based test development. The findings of this study corroborate the literature regarding effective training and professional development in an area so significant to nursing education and student preparation.

Implications of the Findings

The study described the test development practices of nursing faculty in Florida nursing programs. It also identified the differences between ASN and BSN faculty in

utilizing these practices in their test development processes. The implications for nursing education, nursing practice, nursing education research, and public policy are discussed in the following sections.

Implications for Nursing Education

Retention and effective preparation of nursing students are significant concerns for the nursing profession as well as the healthcare community. Effective test development practices contribute to addressing these concerns. The ability to apply evidence-based test development practices is essential to meeting educational objectives. This dissertation study identified and described the testing practices of nursing faculty in pre-licensure programs. Faculty participants' responses to the survey identified areas of test development that can affect the retention and effective assessment of competence in students in pre-licensure nursing programs. The survey focused on test construction, test revision, and beliefs, attitudes, and values of faculty regarding evidence-based test development practices.

The use of test banks in test development has been a continued issue in nursing education. Issues with the validity and reliability of these test items have been discussed significantly in the literature. Test banks are easily available for faculty as a source of questions and are relied upon based on time constraints and lack of item development skills (Clifton & Schriner, 2010; Lampe & Tsaouse, 2010). This has been substantiated by the participant responses in this study. Identification of biases has also been discussed regarding effective test development. This study identified that evaluating tests for bias, both cultural and linguistic, was not considered an important practice by the majority of faculty respondents. These biases in tests affect both the strong students as well as the

weaker students, which leads to poor retention and ultimately affects successful program completion.

It is also important to note the lack of test development training acknowledged by the faculty respondents. The test development and item writing process is difficult and time consuming. Faculty should be required to be trained in the evidence-based practices in order to develop tests that produce valid and reliable assessment data. Based on the faculty responses in the survey, the majority of faculty have had very little to no training in effective test development. It is no wonder that faculty have placed a heavy reliance on test banks and review books to develop their course tests.

The results of this study solidify the supposition that it is essential that nurse educators focus on improving the test development processes by implementing evidence-based practices to produce valid and reliable tests within their programs. This will contribute to increasing the retention of qualified students and ensuring they are effectively prepared as competent and effective nurses for practice.

Implications for Nursing Practice

Retaining qualified students in nursing programs is a significant issue facing many nursing schools nationwide. Effectively preparing these nursing students for clinical practice is also a concern in order to meet the needs of a diverse community. Valid assessment of knowledge and skills is essential in the competency and safety of graduate nurses. This study identifies the attitudes and current practice of nurse educators in the assessment process, specifically in test development. The data collected illustrated that many faculty in nursing education in Florida are not consistently implementing and applying best practices to the test development process. Not using best practices in test

development leads to invalid and inaccurate assessment data on students' knowledge and abilities. This ultimately affects nursing practice performance and the profession's ability to meet the needs of the healthcare community.

Best practices such as assessing items for cultural and linguistic biases contribute to the increased retention of the diverse nursing student population. This diversity in nursing practice is essential in today's multicultural and multi-linguistic healthcare environment, especially with the demographics in Florida. Inaccurate construction and revision in test construction are problematic for nursing practice as they affect the number of nurses graduating from nursing programs which leads to insufficient graduates to fill the open RN positions throughout the nation. In addition, these graduates are expected to be capable of applying information acquired throughout their nursing programs to complicated clinical situations.

Nursing is characterized as a practice profession with significant knowledge and application abilities to ensure competency. Assessing nursing students effectively will ensure these graduate nurses are capable of applying advanced nursing concepts in complex clinical situations (Aiken et al., 2011). The major goal of implementing valid evidence-based tests is to assess student learning outcomes as they relate to discipline-specific guidelines in preparing these students to provide safe and competent care. Therefore, the research data obtained in this study support the continued concern regarding the effective preparation of nursing graduates to function in the dynamic healthcare industry.

Implications for Nursing Research

Evidence-based research is essential for the continued transformation of nursing education. Identifying evidence-based practices in nursing education contributes to the continued preparation of nurses for practice. This research study was guided by Rokeach's beliefs, attitudes, and values theory which addresses the underlying factors that lead to changes in behavior and practices. The *Beliefs, Attitudes, and Values* section of the survey developed by this investigator was based on the premise that understanding these perceptions of nurse educators may lead to changes in testing practices within nursing education. With the identification of current beliefs related to test development through research, changes in poor practices in nursing education can be addressed. The importance of proper testing in nursing education has been documented in the literature as a challenge in preparing nurses for practice. This dissertation study documents and supports previous studies related to poor testing practices in nursing education programs.

Another implication for nursing research is the limited nursing education instruments specifically as it relates to beliefs, attitudes and values of nursing faculty. Utilization of effective instruments will yield rich data which is reliable and valid. However, if there is low faculty participation, this affects the instrument's credibility. Increasing the development and availability of nursing research instruments will enhance and produce strong research in nursing education.

Implications for Public Policy

One of the high priority issues in health care currently is increasing the diversity of the registered nurse population (American Association of Colleges of Nursing, 2015; Ayoola, 2013; Sitzman, 2007). The number of culturally and ethnically diverse students

in nursing programs needs to substantially reflect the diverse communities they serve. However, there have been significant discussions about the poor retention and increased attrition rates of minority nursing students in programs across the U.S. The National League for Nursing (2013) reported that the low number of males and racial-ethnic minority students graduating from nursing programs continues to be a significant issue that needs to be addressed.

There are different reasons for this disparity, one of which is the testing practices in nursing programs. Unfair and invalid testing practices affect minority students on a larger scale than non-minority students, which presents an ethical issue in nursing education. Often students do not succeed in nursing courses because of challenges with testing and biases with construct validity. This dissertation study supports the literature that identifies that test development and item development practices contribute to increased attrition of these students. The faculty participants in this study reinforced the premise that evidence-based practices in test development are not consistently being done. This leads to invalid and bias tests which affect all students, but especially minority students, and which ultimately affects the successful completion of these nursing programs. In order to retain a diverse nursing student population, evidence-based test development must be a priority for all pre-licensure nursing programs. Assurance of fair testing in pre-licensure nursing programs potentially will enhance the diversity of the nursing workforce by removing a significant barrier to student success.

Limitations

The limitations of a study need to be identified and discussed in all research studies (Polit & Beck, 2012). In this dissertation study there were several limitations

noted. The most significant limitation in this study was the sample size. Although 95

deans and directors of pre-licensure programs in Florida were sent emails regarding the

study, only 117 faculty members participated in the study. Of the 117 surveys obtained,

the final number of usable surveys after data cleaning was 104, which is approximately

50% of the anticipated and calculated sample size needed. This number of participants

could be related to the gatekeeper method used to disseminate the survey instrument. In

addition, according to the Florida's Workforce Supply Characteristics and Trends Survey

(2018) there was an estimated 2,410 Registered Nurses whose primary practice is in the

academic setting. However, there is no data specifying how many were employed in

ASN, BSN, or MSN programs. Estimation of response rate and true faculty population is

difficult to calculate. Another limitation to this study, is the length and reliability of the

survey instrument. Although pretesting of the survey was completed with 15 nurse

educators, the investigator was concerned about the length of the survey and the fact that

it may be a deterrent to its completion. The reliability of the instrument subscales were

not as conclusive as the other reliability and validity measures demonstrated in the

original instrument by Killingsworth (2013). The Cronbach's alpha was below the

benchmark of ≥ 0.70 which could be related to the limited sample size, demographics of

the participants and the utilization of two subscales instead of the entire instrument for

analysis.

The final limitation of the study is the faculty understanding of the content of the

survey. The understanding of evidence-based practices is very subjective and participants

may have thought the processes they were implementing were best practices. The faculty

may also have answered the items in ways that reflected positively on them as educators.

Evidence-based practice is a major focus of nursing education, so it may have been a desire of the nurse educators to answer according to the concept of evidence-based practices instead of truly implementing them in their test development practices.

Recommendations for Future Research

The use of evidence-based test development practices is a significant issue in nursing education. Student progression and success are directly affected by poorly developed tests and have significant consequences to nursing education and nursing practice. This study sought to identify the test development practices utilized by nursing faculty and their beliefs regarding best practices. However, more investigation into this issue is needed. The first recommendation for future research is to utilize a qualitative approach to obtaining information regarding beliefs, attitudes, and values of nursing faculty as they relate to evidence-based practices. This will add more clarity into the issues that are barriers to using these practices in test development within their courses.

Another recommendation for future research is to study the understanding of evidence-based test development practices. It would be interesting to uncover what faculty consider evidence-based test development practices. It has been this researcher's experience that most faculty do not fully understand concepts such as cultural and linguistic bias in testing. Identifying their true understanding of these concepts within evidence-based practices may lead to discovering why these practices are not being routinely implemented in their test construction processes.

Chapter Summary

Within this chapter, the investigator discussed and evaluated the results of the research study. The results of the survey indicated that faculty do not routinely implement

evidence-based practices in their test development processes, a result which supports previous literature. The major practices and beliefs of faculty related to test development were discussed. The timing, length, and sample size were identified as limitations of the study and may have affected the results. Literature was integrated and discussed as it related to the study data. The implications of best practices in test development on nursing education, nursing research, nursing practice, and public policy were discussed. Future research topics were suggested and types of methodology recommended. Nursing education needs to continue to advocate for fair and valid testing in pre-licensure programs in order to prepare competent and safe nurses for practice.

References

Abate, M. A., Stamatakis, M. K., & Haggett, R. R. (2003). Excellence in curriculum development and assessment. *American Journal of Pharmaceutical Education*, *67*(3), 1-22.

Abedi, J. (2006). Language issues in item development. In S. M. Downing & T. M. Haladyna, (Eds.), *Handbook of test development* (pp. 287-301). Mahwah, NJ: Lawrence Erlbaum Associates.

Accreditation Commission for Education in Nursing (ACEN). (2013). *Accreditation manual: 2013 standards and criteria*. Atlanta, GA: Author.

Accreditation Commission for Education in Nursing (ACEN). (2017). *Accreditation Manual: 2017 standards and criteria*. Atlanta, GA: Author.

Aiken, L. H., Cimiotti, J. P., Sloane, D. M., Smith, H. L., Flynn, L., & Neff, D. F. (2011). The effects of nurse staffing and nurse education on patient deaths in hospitals with different nurse work environments. *Medical Care*, *49*(12), 1047-1053.

Amaro, D. J., Abriam-Yago, K., & Yoder, M. (2006). Perceived barriers for ethnically diverse students in nursing programs. *Journal of Nursing Education, 45*(7), 247-254.

American Association of Colleges of Nursing (AACN). (2008). *The essentials of baccalaureate education for professional nursing practice*. Retrieved from http://www.aacnnursing.org/Portals/42/Publications/BaccEssentials08.pdf

American Association of Colleges of Nursing (AACN). (2011). *Leading initiatives*. Retrieved from http://www.aacn.nche.edu/leading_initiatives_news/news/2011/employment11

American Association of Colleges of Nursing (AACN). (2015). *Leading initiatives*.

Retrieved from https://www.aacnnursing.org/News-Information/Fact-

Sheets/Enhancing-Diversity

American Association of Colleges of Nursing (AACN). (2017). *Fact sheet: Nursing*

shortage. Retrieved from www.aacnnursing.org/news-information/nursing-

shortage.

Anderson, J. K. (2009). The work-role transition of expert clinician to novice academic

educator. *Journal of Nursing Education*, *48*(4), 203-208.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and*

assessing: A revision of Bloom's taxonomy of educational objectives. New York,

NY: Longmans, Green.

Andrews, M., & Roberts, D. (2003). Supporting student nurses learning in and through

clinical practice: The role of the clinical guide. *Nurse Education Today, 23*(7),

474-481.

Aucoin, J. W. (2005). Writing better test items. *Journal for Nurses in Staff Development,*

21(1), 6-9.

Aucoin, J. W. (2008). Achievement tests. In B. Penn (Ed.), *Mastering the teaching role:*

A guide for nurse educators. Philadelphia, PA: FA Davis.

Aucoin, J., & Rodgers, S. (2000). When to take the NCLEX-RN. *Advance for Nurses*,

2(19), 17-18.

Ayoola, A. (2013). Why diversity in the nursing workforce matters. *RWJF News*.

Retrieved from https://www.rwjf.org/en/blog/2013/05/why_diversity_inthe.html

Baker, E. L., Aschbacher, P. R., Niemi, D., & Sato, E. (1992). *CRESST performance assessment models: Assessing content area explanations*. Retrieved from https://cresst.org/wp-content/uploads/R652.pdf

Bednarz, H., Schim, S., & Doorenbos, A. (2010). Cultural diversity in nursing education: Perils, pitfalls, and pearls. *Journal of Nursing Education, 49*(5), 253-260.

Benner, P. (2012). Educating nurses: A call for radical transformation—How far have we come? *Journal of Nursing Education, 51*(4), 183-184.

Benner, P., Sutphen, M., Leonard, V., & Day, L. (2010). *Educating nurses: A call for radical transformation*. San Francisco, CA: Jossey-Bass.

Billings, D. M., & Halstead, J. A. (2011).*Teaching in nursing: A guide for faculty* (4th ed.). St. Louis, MO: Saunders Elsevier.

Bloom, B. S. (1956). Taxonomy of educational objectives. *Handbook I: The cognitive domain*. New York, NY: David McKay Co Inc.

Bosher, S. D. (2003). Barriers to creating a more culturally diverse nursing profession: Linguistic bias in multiple-choice nursing exams. *Nursing Education Perspectives, 24*(1), 25-34.

Bosher, S. D. (2009). Removing language as a barrier to success on multiple-choice nursing exams. In S. D. Bosher & M. D. Pharris (Eds.), *Transforming nursing education: The culturally inclusive environment* (pp. 259-284). New York, NY: Springer.

Bosher, S. D., & Bowles, M. (2008). The effects of linguistic modifications on ESL students' comprehension of nursing course test items. *Nursing Education Perspectives, 29*(3), 165-172.

Bosher, S. D., & Pharris, M. D. (2009). *Transforming nursing education: The culturally inclusive environment*. New York, NY: Springer Publishing.

Budd, J. (2001). *Knowledge and knowing in library and information science: A philosophical framework*. Lanham, MD: Scarecrow Press.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items* (Vol. 4). Thousand Oaks, CA: Sage.

Carrick, J. (2011). Student achievement and NCLEX-RN success: Problems that persist. *Nursing Education Perspectives, 32*(2), 78-83.

Cascoe, K., Stanley, S., Stennett, R., & Allen, C. (2017). Undergraduate nursing students at risk of failure. *Nurse Education Today*, *52*, 121-122.

Clarke, S. (2017). What you need to know about the NCLEX-RN. *Nursing Management, 48*(10), 21-23.

Clark-Goff, K., & Eslami, Z. (2016). Exploring change in preservice teacher's beliefs about English language learning and teaching. *Iranian Journal of Language Teaching Research, 4*(3), 21-36.

Clifton, S. L., & Schriner, C. L. (2010). Assessing the quality of multiple-choice test items. *Nurse Educator, 35*(1), 12-16.

Commission on Collegiate Nursing Education (CCNE). (2017). *Standards for accreditation of baccalaureate and graduate degree nursing programs*. Retrieved from http://www.aacnnursing.org/Portals/42/CCNE/PDF/Procedures.pdf

Commission on Collegiate Nursing Education (CCNE). (2018). *Standards for accreditation of baccalaureate and graduate degree nursing programs*. Retrieved

from https://www.aacnnursing.org/CCNE-Accreditation/Resource-

Documents/CCNE-Standards-Professional-Nursing-Guidelines

Center for Research on Evaluation, Standards, and Student Testing (CRESST). (2018). *Is this a trick question? A short guide to writing effective test questions.* Retrieved from https://cresst.org/education.

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. Thousand Oaks, CA: Sage Publications.

Creswell, J. W., & Clark, V. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Cross, K. J. (2000). *Cognitive levels of multiple-choice items on teacher-made tests in nursing education*. (Doctoral Dissertation). Retrieved from Proquest.

DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *The Canadian Journal for the Scholarship of Teaching and Learning, 2*(2), 1-23.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York, NY: John Wiley and Sons.

Downing, S. M. (2004). Reliability: On reproducibility assessment data. *Medical Education, 38*(9), 1006-1012.

Downing, S. M. (2005). The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement tests in medical education. *Advances in Health Sciences, 10*, 133-143.

Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 3-24). Mahwah, NJ: Lawrence Erlbaum Associates.

Downing, S. M., & Haladyna, T. M. (2011). *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dudas, K. I. (2014). *Living in two worlds: Experiences of non-native English speakers in an accelerated second-degree baccalaureate nursing program* (Order No. 3640016). Available from Dissertations & Theses @ Nova Southeastern University; ProQuest Central; ProQuest Dissertations & Theses Global. (1620832617).

Duphily, N. H. (2011). The experience of novice nurse faculty in an associate degree education program. *Teaching and Learning in Nursing*, *6*(3), 124-130.

Duty, S. M., Christian, L., Loftus, J., & Zappi, V. (2016). Is cognitive test-taking anxiety associated with academic performance among nursing students? *Nurse Educator, 41*(2), 70-74.

Eckardt, R. (1970). Beliefs, attitudes and values: A theory of organization and change by Milton Rokeach—Review. *Journal for the Scientific Study of Religion, 9*(3), 268-269.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks, CA: Sage.

Fiset, V. J., Graham, I. D., & Davies, B. L. (2017). Evidence-based practice in clinical nursing education: A scoping review. *Journal of Nursing Education, 56*(9), 534-541.

Florida Board of Nursing. (2018). *Licensure by examination FAQs*. Retrieved from

http://floridasnursing.gov/nursing-faqs/licensure-by-examination/

Fricker, R. D., & Schonlau, M. (2002). Advantages and disadvantages of Internet

research surveys: Evidence from the literature. *Field Methods*, *14*(4), 347-367.

Gorsuch, R. (1969). Beliefs, attitudes and values, by Milton Rokeach – Review.

American Sociological Review, 34(2), 267-268.

Griffiths, M. J., Papastrat, K., Czekanski, K., & Hagan, K. (2004). The lived experience

of NCLEX failures. *Journal of Nursing Education, 43*(7), 322-325.

Haladyna, T. M., & Downing, S. M. (1985, April). *A quantitative review of research on

multiple-choice item writing*. Paper presented at the Annual Meeting of the

American Educational Research Association, Chicago, IL.

Haladyna, T. M., & Downing, S. M. (1989a). A taxonomy of multiple choice item-

writing rules. *Applied Measurement in Education, 2*(1), 37-50.

Haladyna, T. M., & Downing, S. M. (1989b). Validity of a taxonomy of multiple-choice

item-writing rules. *Applied Measurement in Education, 2*(1), 51-78.

Halstead, J. A. (2011). The realist adjusts the sails: A commitment to transform nursing

education models. *Nursing Education Perspectives, 32*(6), 357.

Hamdy, H. (2007). Blueprinting in medical education. *New England Journal of Medicine,

3*(56), 387-395.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009).

Research electronic data capture (REDCap)—A metadata-driven methodology

and workflow process for providing translational research informatics support.

Journal of Biomedical Informatics, *42*(2), 377-381.

Hart, A. M., Brennan, C. W., Sym, D., & Larson, E. (2009). The impact of personalized prenotification on response rates to an electronic survey. *Western Journal of Nursing Research*, *31*(1), 17-23.

Hicks, N. A. (2011). Guidelines for identifying and revising culturally biased multiple-choice nursing examination items. *Nurse Educator, 36*, 266-270.

Kantar, L. D. (2014). Assessment and instruction to promote higher order thinking in nursing students. *Nurse Education Today, 34*(5), 789-794

Kantor, S. A. (2010). Pedagogical change in nursing education: One instructor's experience. *Journal of Nursing Education, 49*(7), 414-417.

Killingsworth, E. (2013). *Nursing faculty decision making about best practices in test construction, item analysis, and revision*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3577368)

Killingsworth, E., Kimble, L., & Sudia, T. (2015). What goes into a decision? How nursing faculty decide which best practices to use for classroom testing. *Nursing Education Perspectives, 36*(4), 220-225.

Klisch, M. L. (1994). Guidelines for reducing bias in nursing examinations. *Nurse Educator, 19*(2), 35-39.

Krathwohl, D. (2010). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212-218.

Kreueger, L. (2014). Academic dishonesty among nursing students. *Journal of Nursing Education, 53*, 77-87.

Lampe, S., & Tsaouse, B. (2010). Linguistic bias in multiple choice test questions. *Creative Nursing, 16*(2), 63-67.

Long, K. A. (2003). The Institute of Medicine report: Health professions education: A

   bridge to quality. *Policy, Politics, & Nursing Practice*, *4*(4), 259-262.

Lujan, J. (2007). Linguistic and cultural adaptation needs of Mexican-American nursing

   students related to multiple-choice testing. *Journal of Nursing Education*, *47*(7),

   327-330.

Masters, J. C., Hulsmeyer, B. S., Pike, M. E., Leichty, K., Miller, M. T., & Verst, A. L.

   (2001). Assessment of multiple choice questions in selected test banks

   accompanying textbooks used in nursing education. *Journal of Nursing

   Education, 40*(1), 25-31.

McCabe, J. (2009). Racial and gender microaggressions on a predominantly-White

   campus: Experiences of Black, Latina/o and White undergraduates. *Race, Gender

   & Class, 16*(1/2), 133-151.

Mehrdad, N., Joolee, S., Joulaee, A., & Bahrani, N. (2012). Nursing faculties' knowledge

   and attitude on evidence based practice. *Iranian Journal of Nursing and

   Midwifery Research, 17*(7), 506-511.

Mertens, D. M. (2014). *Research and evaluation in education and psychology:

   Integrating diversity with quantitative, qualitative, and mixed methods*. Thousand

   Oaks, CA: Sage Publications.

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

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

Morrison, S., Nibert, A., & Flick, J. (2006). *Critical thinking and test item writing* (2nd

   ed.). Houston, TX: Health Education Systems.

Morse, J. M., & Niehaus, L. (2009). *Mixed method design: Principles and procedures*. Walnut Creek, CA: Left Coast Press.

Nadeem, N., van der Vleuten, C., & Alfaris, E. A. (2012). Faculty development on item writing substantially improves item quality. *Advances in Health Sciences Education, 17*(3), 369-376.

National Council of State Boards of Nursing (NCSBN). (2006). *Assessing the impact of English as a second language on licensure examination*. Retrieved from https://www.ncsbn.org/ESL_Licensure.pdf

National Council of State Boards of Nursing (NCSBN). (2008). *Nursing faculty qualifications and roles*. Retrieved from https://www.ncsbn.org/Final_08_Faculty _Qual_Report.pdf

National Council of State Boards of Nursing (NCSBN). (2016a). *Exam psychometrics*. Retrieved from https://www.ncsbn.org/1230.htm

National Council of State Boards of Nursing (NCSBN). (2016b). *NCLEX-RN test plan*. Retrieved from https://www.ncsbn.org/RN_Test_Plan_2016_Final.pdf

National Council of State Boards of Nursing (NCSBN). (2017). *NCLEX examinations*. Retrieved from https://www.ncsbn.org/nclex.htm

National League for Nursing (NLN). (2005). *Core competencies of nurse educators with task statements*. Retrieved from www.nln.org

National League for Nursing. (2009). *A commitment to diversity in nursing and nursing education. Reflections and dialogue*. http://www.nln.org/about/position-statements/nln-reflections-dialogue/read/dialoguereflection/2009/01/02/reflection-dialogue-3-

National League for Nursing (NLN). (2011). *Certification for nurse educators (CNE).*

    Retrieved from http://www.nln.org/professional-development-

    programs/Certification-for-Nurse-Educators

National League for Nursing (NLN). (2012a). *Nurse educator core competencies.*

    Retrieved from http://www.nln.org/professional-development-

    programs/competencies-for-nursing-education/nurse-educator-core-competency

National League for Nursing (NLN). (2012b). *The fair test imperative in nursing*

    education. Retrieved for http://www.nln.org/aboutnln/livingdocuments/pdf

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum*

    Studies, 19(4), 317-328.

Nick, J., Sharts-Hopko, N. C., & Leners, D. W. (2013). From committee to commission:

    The history of the NLN's academic certified nurse educator program. *Nursing*

    Education Perspectives, 34(5), 298-302.

Nilbert, A., & Morrison, S. (2013). HESI testing: A history of evidence-based research.

    Journal of Professional Nursing, 29, S2-S4.

Notar, C. E., Zuelke, D. C., Wilson, J. D., & Yunker, B. D. (2004). The table of

    specifications: Insuring accountability in teacher made tests. *Journal of*

    Instructional Psychology, 31(2), 115-129.

Oermann, M. H., & Gaberson, K. B. (2016). *Evaluation and testing in nursing education*

    (5th ed.). New York, NY: Springer Publishing.

Oermann, M. H., Saewert, K. J., Charasika, M., & Yarbrough, S. S. (2009). Assessment

    and grading practices in schools of nursing: National survey findings part

    I. *Nursing Education Perspectives*, *30*(5), 274-278.

Ortelli, T. A. (2006). Defining the professional responsibilities of academic nurse educators: The results of a national practice analysis. *Nursing Education Perspectives, 27*(5), 242-246.

Pallant, J. (2010) *SPSS survival manual: A step by step guide to data analysis using SPSS*. (4th Ed.). Buckingham, UK: Open University Press.

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham, MD: Rowman & Littlefield.

Phillips, J. M., & Malone, B. (2014). Increasing racial/ethnic diversity in nursing to reduce health disparities and achieve health equity. *Public Health*, *129*(12), 45-50.

Poindexter, K., Hagler, D., & Lindell, D. (2015). Designing authentic assessment: Strategies for nurse educators. *Nurse Educator, 40*(1), 36-40.

Polit, D. F., & Beck, C. T. (2008). *Nursing research: Generating and assessing evidence for nursing practice* (8th ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.

Polit, D. F., & Beck, C. T. (2012). *Nursing research: Generating and assessing evidence for nursing practice* (9th ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice, 48*(4), 4-22.

Powers, B. A., & Knapp, T. R. (1995). *A dictionary of nursing theory and research*. Thousand Oaks, CA: Sage.

Punch, K. (2005). *Introduction to social research quantitative and qualitative approaches*. Thousand Oaks, CA: Sage.

Raines, D. A. (1992). *An analysis of the values influencing neonatal nurses' perceptions and behaviors in selected ethical dilemmas* (Dissertation).

Randolph, P. K. (2017). Standardized testing practices: Effect on graduation and NCLEX pass rates. *Journal of Professional Nursing, 33*(3), 224-228.

Richards, E., & Stone, C. (2008). Student evaluation of a standardized comprehensive testing program. *Nursing Education Perspectives, 29*(6), 363-365.

Richardson, V. (1996).The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), *Handbook of research on teacher education* (2nd ed., pp. 102-119). New York, NY: Macmillan.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24*(2), 3-13.

Rokeach, M. (1968). A theory of organization and change within value-attitude systems. *Journal of Social Issues, 24*(1), 13-33.

Samdahl, D. M. (1999). Epistemological and methodological issues in leisure research. In T. L. Burton & E. L. Jackson (Eds.), *Leisure studies: Prospects for the twenty-first century* (pp. 119-133). State College, PA: Venture Publishing.

Sanders, B. (2018, September). *NCLEX overview: Psychometrics and test development.* Paper session presented at the National Council of State Boards of Nursing NCLEX Conference, Charlotte, NC.

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, *44*(4), 409-432.

Schmeiser, C. B., & Welch, C. J. (2006). Test development. *Educational measurement*, *4*, 307-353.

Schmidt, B., & MacWilliams, B. (2011). Admission criteria for undergraduate nursing

    programs: A systematic review. *Nurse Educator, 36*(4), 171-174.

Scotland, J. (2012). Exploring the philosophical underpinnings of research: Relating

    ontology and epistemology to the methodology and methods of the scientific,

    interpretive, and critical research paradigms. *English Language Teaching*, *5*(9), 9.

Sereno, K. (2012). Explaining theories of persuasion, communication and social science.

    Lecture Notes, 32, 22-24.

Siegel, T. (2015). *Assessment practices at an associate degree nursing program*

    (Doctoral dissertation). Retrieved from https://scholarworks.waldenu.edu/cgi/

    viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1602&c

    ontext=dissertations

Sitzman, K. L. (2007). Diversity and the NCLEX-RN: A double-loop approach. *Journal*

    of Transcultural Nursing, *18*(3), 271-276.

Smith, J. K. (1983). Quantitative versus qualitative research: An attempt to clarify the

    issue. *Education Researcher, 12*(3), 6-13.

Su, W. M., Osisek, P. J., Montgomery, C., & Pellar, S. (2009). Designing multiple-choice

    test items at higher cognitive levels. *Nurse Educator, 34*, 223-227.

Suskie, L. (2009). Assessing student learning: A common sense guide (2nd ed.). San

    Francisco, CA: Jossey-Bass.

Sutherland, K., Schwartz, J., & Dickison, P. (2012). Best practices for writing test items.

    Journal of Nursing Regulation, 3(2), 35-39.

Tarrant, M., Knierim, A., Hayes, S., & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education Today, 26*(8), 662-671.

Tarrant, M., & Ware, J. (2008). Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Medical Education, 42*(2), 198-206.

Tarrant, M., & Ware, J. (2011). Guidelines for developing high quality multiple-choice assessments. *Medical Education, 45*(3), 1-15.

Tarrant, M., & Ware, J. (2012). A framework for improving the quality of multiple-choice assessments. *Nurse Educator*, *37*(3), 98-104.

Taxis, J. (2002). The underrepresentation of Hispanics/Latinos in nursing education: A deafening silence. *Research and Theory for Nursing Practice, 16*(4), 249-262.

Theart, C. J., & Smit, I. (2012). The status of academic integrity amongst nursing students at a nursing education institution in the Western Cape. *Curationis*, *35*(1), 1-8.

Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLoS Medicine*, *2*(10), e267.

Victor, B., & Cullen, J. B. (1988). The organizational bases of ethical work climates. *Administrative Science Quarterly, 33*(1), 101-125.

Walloch, J. A. (2006). *Assessment practices in the nursing classroom: An exploration of educators' assessment of students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database.

Wendt, A., & Kenny, L. (2009). Alternative item types: Continuing the quest for authentic testing. *Journal of Nursing Education*, *48*, 150-156.

Whorley, E. (2018). *The implementation of the essential competencies for evidence-based practice in baccalaureate nursing education* (Doctoral dissertation).

Yeom, Y. (2013). An investigation of predictors of NCLEX-RN outcomes among content standardized tests. *Nurse Education Today*, *33*(12), 1523-1528.

Zaidi, N. L. B., Grob, K. L., Monrad, S. M., Kurtz, J. B., Tai, A., Ahmed, A. Z., . . . & Santen, S. A. (2018). Pushing critical thinking skills with multiple-choice questions: Does Bloom's taxonomy work? *Academic Medicine*, *93*(6), 856-859.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, *16*(4), 323-342.

Appendix A

Cover Letter to the Deans and Directors

Dear Nursing Education Leader,

My name is Richild Berrick and I am a doctoral student at Nova Southeastern University in Fort Lauderdale, Florida. For my dissertation, I am examining faculty beliefs, attitudes, and values as it relates to the use of evidence-based practices in test development. Because you are a nurse administrator of a pre-licensure RN program, I am writing to ask whether you would be willing to pass along this survey information to your faculty who may be interested in participating in this research study. You are under no obligation to share this information with your nurse educators. The survey will require approximately 30 minutes to complete. There is no compensation for responding nor is there any known risk. Be assured that all information will remain confidential.

Thank you for taking the time to assist me in my academic endeavors. The data collected will provide useful information regarding test development practices in nursing education. If you require additional information or have questions, please contact me at the number listed below.

Thank you for your assistance in this important endeavor,

Sincerely,

Richild Berrick PhD(c), MSN/Ed, RN, CNE
Principal Investigator   Dissertation Chair
Nova Southeastern University   Nova Southeastern University

Appendix B

Participant Letter of Consent

NSU NOVA SOUTHEASTERN UNIVERSITY
Ron and Kathy Assaf College of Nursing

Participant Letter for Anonymous Surveys
NSU Consent to be in a Research Study Entitled
The Use of evidence Based Test Development in Pre-Licensure Nursing Programs: A
Descriptive Study of Faculty Beliefs, Attitudes, and Values.

Who is doing this research study?

This person doing this study is Richild Berrick with the Ron and Kathy Assaf College of Nursing.
They will be helped by Julia Aucoin, Dissertation Chair.

Why are you asking me to be in this research study?

You are being asked to take part in this research study because you are a nursing faculty who
actively participates in test development and item writing within your Florida pre-licensure
nursing program.

Why is this research being done?

The purpose of this study is to find out the beliefs, attitudes, and values of nursing faculty
regarding the utilization of test development best practices in undergraduate nursing programs.

What will I be doing if I agree to be in this research study?

You will be taking a one-time, anonymous survey. The survey will take approximately 30
minutes to complete.

Are there possible risks and discomforts to me?

This research study involves minimal risk to you. To the best of our knowledge, the things you
will be doing have no more risk of harm than you would have in everyday life.

What happens if I do not want to be in this research study?

You can decide not to participate in this research and it will not be held against you. You can
exit the survey at any time.

Will it cost me anything? Will I get paid for being in the study?

There is no cost for participation in this study. Participation is voluntary and no payment will be
provided.

How will you keep my information private?

Your responses are anonymous. Information we learn about you in this research study will be handled in a confidential manner, within the limits of the law. The information collected from the survey will be stored in an online survey platform and be accessible only by the login and password of the researcher. This data will be available to the researcher, the Institutional Review Board and other representatives of this institution, and any granting agencies (if applicable). All confidential data will be kept securely Data will be stored in the Research Electronic Data Capture (REDCap) online survey database and will be password protected. All data will be kept for 36 months from the end of the study and destroyed after that time by deleting all information form the program.

Who can I talk to about the study?

If you have questions, you can contact Richild Berrick at ███████████ and ███████████ that will be readily available during and after normal work hours.

███

Do you understand and do you want to be in the study?

If you have read the above information and voluntarily wish to participate in this research study, please click on the survey link provided in this email.

Appendix C

IRB Approval Letter

NSU NOVA SOUTHEASTERN UNIVERSITY
Institutional Review Board

MEMORANDUM

To: **Richild Berrick**

From: ████████████████

        Center Representative, Institutional Review Board

Date: **November 5, 2018**

Re: **IRB #: 2018-567; Title, "The Use of evidence Based Test Development in Pre-Licensure Nursing Programs: A Descriptive Study of Faculty Beliefs, Attitudes, and Values."**

I have reviewed the above-referenced research protocol at the center level. Based on the information provided, I have determined that this study is exempt from further IRB review under **45 CFR 46.101(b)** (**Exempt 2: Interviews, surveys, focus groups, observations of public behavior, and other similar methodologies**). You may proceed with your study as described to the IRB. As principal investigator, you must adhere to the following requirements:

1)   CONSENT: If recruitment procedures include consent forms, they must be obtained in such a manner that they are clearly understood by the subjects and the process affords subjects the opportunity to ask questions, obtain detailed answers from those directly involved in the research, and have sufficient time to consider their participation after they have been provided this information. The subjects must be given a copy of the signed consent document, and a copy must be placed in a secure file separate from de-identified participant information. Record of informed consent must be retained for a minimum of three years from the conclusion of the study.

2)   ADVERSE EVENTS/UNANTICIPATED PROBLEMS: The principal investigator is required to notify the IRB chair and me ████████████████████████████████ of any adverse reactions or unanticipated events that may develop as a result of this study. Reactions or events may include, but are not limited to, injury, depression as a result of participation in the study, life-threatening situation, death, or loss of confidentiality/anonymity of subject. Approval may be withdrawn if the problem is serious.

3)   AMENDMENTS: Any changes in the study (e.g., procedures, number or types of subjects, consent forms, investigators, etc.) must be approved by the IRB prior to implementation. Please be advised that changes in a study may require further review depending on the nature of the change. Please contact me with any questions regarding amendments or changes to your study.

The NSU IRB is in compliance with the requirements for the protection of human subjects prescribed in Part 46 of Title 45 of the Code of Federal Regulations (45 CFR 46) revised June 18, 1991.

Appendix D

Survey Instrument

Faculty Beliefs, Attitudes and Values

Please complete the survey below.

Thank you!

Welcome to the Faculty Beliefs, Attitudes and Values in Evidence-Based Test Development research study. The following questions are to identify eligibility to participate in this study.

Are you a nursing faculty member teaching in a pre-licensure program?
○ Yes
○ No

Do you currently develop tests in at least one nursing course in your program?
○ Yes
○ No

In this section, the questions are intended to collect information about you, your teaching experience, and the nursing education program you work in. Please indicate the option that best describes you and your nursing education program.

What type of pre-licensure nursing program do you teach in?
○ ASN
○ BSN
○ Both

What Florida area code is your nursing program in? (add the first 3 digits of the area code)
_____

How many years have you been writing test items?
○ 1-3
○ 4-7
○ 8-10
○ 10-15
○ 16 years or more

What is the method of test delivery for classroom tests in the identified course?
○ Paper and pencil
○ Computer
○ Both

What is your highest degree completed?
○ BSN
○ MSN
○ Doctorate

If you have a master's degree, what role did your education formally prepare you for?
○ Nurse Educator
○ Nurse Administrator
○ Nurse Practitioner
○ Certified Nurse Midwife
○ Certified Registered Nurse Anesthetist
○ Clinical Nurse Specialist
○ Clinical Nurse Leader
○ None of the above
○ I do not have a Master's in nursing

Please indicate the amount of course work you have received in test development.
○ No course in test development
○ One course in test development
○ More than one course in test development

Have you ever participated in a professional development program focusing on test development?
○ Yes
○ No
○ I cannot remember

Do you hold certification as a nurse educator? (e.g.
CNE, NPD)

- ◯ Yes
- ◯ No

In the past five years, have there been any concerns
expressed to faculty from administration about the
NCLEX-RN pass rates for your nursing education
program?

- ◯ Yes
- ◯ No
- ◯ I do not know

Test Construction

The following are components of test construction that faculty can use when constructing a test.
Please indicate how often you use each component when developing test items within the identified
nursing course by selecting from least descriptive to most descriptive.

Course objectives

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

Class or unit objectives

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

Major content topics

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

Specific content topics

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

A test blueprint or table of specifications

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

The NCLEX-RN test plan

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

Peer review of test items

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

Higher cognitive levels according to Bloom's taxonomy (e.g. application, analysis, evaluation)

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

Clinical context for test items.

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

Plausible distractors in multiple-choice test items

◯ Least descriptive ◯ ◯ ◯ ◯ ◯ Most descriptive

Even distribution of correct answer in multiple-choice options.

○ Least descriptive   ○ ○ ○ ○ ○ ○ Most descriptive

Use various test item types (e.g., multiple-choice, select all that apply, fill in the blank, etc.)

○ Least descriptive   ○ ○ ○ ○ ○ ○ Most descriptive

Overall, how skilled are you at test construction?

○ Not Skilled   ○ ○ ○ ○ ○ ○ Very Skilled

The following are potential sources of your items for tests within the identified nursing course. Please indicate the frequency with which you use the sources from Never use to Always use.

Current course faculty member.

○ Never use   ○ ○ ○ ○ ○ ○ Always use

Yourself

○ Never use   ○ ○ ○ ○ ○ ○ Always use

Textbook tests or test banks

○ Never use   ○ ○ ○ ○ ○ ○ Always use

Review book tests or test item banks

○ Never use   ○ ○ ○ ○ ○ ○ Always use

Previous nursing faculty member-made tests

○ Never use   ○ ○ ○ ○ ○ ○ Always use

Nursing education program-developed test item banks

○ Never use   ○ ○ ○ ○ ○ ○ Always use

The following is a list of different information faculty can obtain about test items after a test is administered. Please indicate how often you use the information after test administration for tests within the identified nursing course. Please answer on a scale of "not at all" to "all the time"

The number of students who answered each question incorrectly.

○ Not at all   ○ ○ ○ ○ ○ ○ All the time

The number of students who answered each question correctly (Difficulty level or p-value).

○ Not at all   ○ ○ ○ ○ ○ ○ All the time

A question's ability to discriminate between the high and low scoring students (discrimination index to point biserial coefficient).

○ Not at all　○ ○ ○ ○ ○ ○ All the time

The frequency of distractor choices with each test question.

○ Not at all　○ ○ ○ ○ ○ ○ All the time

The discrimination between the high and low scoring students choosing distractors (distractor discrimination).

○ Not at all　○ ○ ○ ○ ○ ○ All the time

The central tendency (mean, standard deviation) of the student grades on the test.

○ Not at all　○ ○ ○ ○ ○ ○ All the time

Overall, how skilled are you at using test item analyses?

○ Not Skilled　○ ○ ○ ○ ○ ○ Very Skilled

Test Revision

The following are actions faculty can perform when revising classroom tests. Please indicate how often you perform these actions during test revision for tests within the identified nursing course. Please answer on a scale of " Not at all" to "All the time."

Use item analysis data when determining to keep or eliminate test questions before finalizing test scores.

○ Not at all　○ ○ ○ ○ ○ ○ All the time

Compare item analysis data for test questions used repeatedly from one term to another.

○ Not at all　○ ○ ○ ○ ○ ○ All the time

Use distractor discrimination to revise test items.

○ Not at all　○ ○ ○ ○ ○ ○ All the time

Use difficulty level of test items to revise test items.

○ Not at all　○ ○ ○ ○ ○ ○ All the time

Assess for linguistic/cultural bias in test items.

○ Not at all　○ ○ ○ ○ ○ ○ All the time

Assess for changes in domain content based upon new research data.

○ Not at all　○ ○ ○ ○ ○ ○ All the time

Assess for outdated language used in test items.

○ Not at all　○ ○ ○ ○ ○ ○ All the time

Change test items to ensure test security.

○ Not at all   ○ ○ ○ ○ ○ ○ All the time

Change test items to reflect emphasis in classroom content.

○ Not at all   ○ ○ ○ ○ ○ ○ All the time

Change test items to ensure sufficient sampling of content.

○ Not at all   ○ ○ ○ ○ ○ ○ All the time

Overall, how skilled are you at test revision?

○ Not Skilled   ○ ○ ○ ○ ○ ○ Very Skilled

Teacher Beliefs

The following statements pertain to your beliefs related to evaluation of students in the classroom. Please indicate on a scale of " Least descriptive" to "Most descriptive" the extent to which these statements describe your beliefs.

What is best for the students is the major consideration.

○ Least descriptive   ○ ○ ○ ○ ○ ○ Most descriptive

In this nursing education program, nursing faculty members are expected to follow their own personal and moral beliefs during student evaluation.

○ Least descriptive   ○ ○ ○ ○ ○ ○ Most descriptive

During student evaluation, nursing faculty members look out for the student's good.

○ Least descriptive   ○ ○ ○ ○ ○ ○ Most descriptive

Each nursing faculty member in this nursing education program decides for himself or herself what is right or wrong in student evaluations.

○ Least descriptive   ○ ○ ○ ○ ○ ○ Most descriptive

It is important to follow the nursing education program's evaluation rules and procedures during student evaluation.

○ Least descriptive   ○ ○ ○ ○ ○ ○ Most descriptive

The most important concern in this nursing education program during student evaluation is the individual faculty member's sense of right and wrong.

○ Least descriptive   ○ ○ ○ ○ ○ ○ Most descriptive

Successful student evaluation occur by strictly following the nursing education program's evaluation rules and procedures.

○ Least descriptive   ○ ○ ○ ○ ○ ○ Most descriptive

In this nursing program, nursing faculty members are guided by their own personal ethics.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

During student evaluation, it is expected that nursing faculty members will always do what is right for the student.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

Nursing faculty members in this nursing education program adhere to the program's rules and procedures of student evaluation.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

The most important concern in student evaluation is the good of the student body as a whole.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

In student evaluations, the best practices of student evaluation are the major consideration.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

Nursing faculty members in this nursing education program are expected to comply with best practices of student evaluation over and above other considerations.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

In this nursing education program, the most efficient way to move students through the nursing program is the right way for student evaluations.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

In this nursing education program, nursing faculty members are expected to strictly follow best practices of student evaluation.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

In this nursing education program, nursing faculty members are expected to above all move students efficiently through the nursing program.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

In this nursing education program, the first consideration is whether a decision regarding student evaluation violates any best practices of student evaluation.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

The major concern of nursing faculty members in this nursing education program is always what is best for the students.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

All nursing faculty members are expected to adhere to the nursing education program's evaluation rules and procedures.

○ Least descriptive ○ ○ ○ ○ ○ ○ Most descriptive

Beliefs, Attitudes, and Values of Evidence-Based Test Development

In this section, the items pertain to your beliefs, attitudes and values of testing and test development. Please indicate on a scale from "strongly disagree" to "strongly agree" in which these statements describe your beliefs, attitudes and values.

Evidence-based test development is an essential part of my role as a nurse educator.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

I understand the evidence-based practices in test development.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

I always implement evidence-based practices in my test development process.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Knowledge of evidence-based test development is important to me.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

I am confident in my ability to prepare evidence-based test items.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

It is the responsibility of all faculty members to participate in evidence-based test development.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

The more time I spend developing evidence-based test items, the less time is available for other critical educational duties.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Using test bank items saves me time in the test development process.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Test bank items are just as effective as faculty-developed items.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Standardized tests identify the students that should/should not progress in the nursing program.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Test development using evidence-based practices, effectively prepares students for NCLEX-RN.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Test development using evidence-based practices effectively prepares students for standardized tests (e.g. HESI, ATI, etc.).

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Evidence-based test development is a skill that all nursing faculty must have.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Faculty-made tests are not as important as standardized tests (e.g. HESI, ATI, etc.) in preparing for the NCLEX-RN.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

I place more emphasis on evidence-based teaching than on evidence based test development.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Evidence-based test development skills are useful but not a necessary component of pre-licensure nursing education.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Most test banks are valid and reliable.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

I often restructure test bank items to change their difficulty.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Always evaluating my tests for linguistic and cultural bias is important in my test development process.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Linguistic and cultural bias is a common problem when using test banks.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

It is important to me to have other faculty review my test prior to administration.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

I value feedback from my colleagues on tests I have developed.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

My nursing program places a priority on evidence-based test development.

○ Strongly disagree ○ ○ ○ ○ ○ ○ Strongly agree

Decision Making

In this section, the items pertain to factors that may influence your decision making about test development. Please indicate how important each of these factors are in influencing your decisions regarding test construction, test item analysis, and test revision.

Tradition and past practice.

○ Minimally important   ○ Moderately important   ○ Very important

Research about effectiveness of various strategies.

○ Minimally important   ○ Moderately important   ○ Very important

Educational soundness or congruence with educational standards.

○ Minimally important   ○ Moderately important   ○ Very important

Peer pressure.

○ Minimally important   ○ Moderately important   ○ Very important

Regulatory standards.

○ Minimally important ○ Moderately important ○ Very important

NCLEX-RN pass rates.

○ Minimally important ○ Moderately important ○ Very important

Time needed to develop test.

○ Minimally important ○ Moderately important ○ Very important

Time needed to implement test.

○ Minimally important ○ Moderately important ○ Very important

Time needed to conduct test item analysis.

○ Minimally important ○ Moderately important ○ Very important

Time needed to revise test items.

○ Minimally important ○ Moderately important ○ Very important

Material and equipment needed for test construction.

○ Minimally important ○ Moderately important ○ Very important

Openness to creativity.

○ Minimally important ○ Moderately important ○ Very important

Incentives (or lack thereof) for innovation.

○ Minimally important ○ Moderately important ○ Very important

Legal defensibility.

○ Minimally important ○ Moderately important ○ Very important

Personal ethics.

○ Minimally important ○ Moderately important ○ Very important

---

Decision Making Processes

People differ in the way they go about making decisions. Please indicate how you make decisions by selecting for each item the responses that best fits your usual style.

When making decisions:

I feel as if I'm under tremendous time pressure when making decisions.

○ Not true ○ Sometimes true ○ True for me

I like to consider all of the alternatives.

○ Not true ○ Sometimes true ○ True for me

I prefer to leave decisions to others.

○ Not true ○ Sometimes true ○ True for me

I try to find out the disadvantages of all alternatives.

○ Not true ○ Sometimes true ○ True for me

I waste a lot of time on trivial matters before getting to the final decisions.

○ Not true ○ Sometimes true ○ True for me

I consider how best to carry out the decision.

○ Not true ○ Sometimes true ○ True for me

Even after I have made a decision I delay acting upon it.

○ Not true ○ Sometimes true ○ True for me

When after making decisions I like to collect lots of information.

○ Not true ○ Sometimes true ○ True for me

I avoid making decisions.

○ Not true ○ Sometimes true ○ True for me

When I have to make a decision I wait a long time before starting to think about it.

○ Not true ○ Sometimes true ○ True for me

I do not like to take responsibility for making decisions.

○ Not true ○ Sometimes true ○ True for me

I try to be clear about my objectives before choosing.

○ Not true ○ Sometimes true ○ True for me

The possibility that small things might go wrong causes me to swing abruptly in my preferences.

○ Not true ○ Sometimes true ○ True for me

If a decision can be made by me or another person I let the other person make it.

○ Not true ○ Sometimes true ○ True for me

Whenever I face a difficult decision I feel pessimistic about finding a good solution.

○ Not true ○ Sometimes true ○ True for me

I take a lot of care before choosing.

○ Not true ○ Sometimes true ○ True for me

I do not make decisions unless I really have to.

○ Not true ○ Sometimes true ○ True for me

I delay making decisions until it's too late.

○ Not true ○ Sometimes true ○ True for me

I prefer that people who are better informed decide for me.

○ Not true   ○ Sometimes true   ○ True for me

After a decision is made I spend a lot of time convincing myself it was correct.

○ Not true   ○ Sometimes true   ○ True for me

I put off making decisions.

○ Not true   ○ Sometimes true   ○ True for me

I cannot think straight if I have to make a decision in a hurry.

○ Not true   ○ Sometimes true   ○ True for me

Appendix E

Permission Letter to Use Original Survey

Appendix F

Original Survey by Killingsworth

101

Screening Tool

Welcome to the Nursing Faculty Decision Making in Classroom Test Construction, Test Item Analysis, and Test Revision. The following question will ascertain if you are eligible to participate in the study. If you are eligible, the informed consent will be presented shortly.

Are you a nursing faculty member teaching in a BSN program?
Yes   No

Do you have a minimum of 2 years of full time teaching experience in a BSN program?
Yes   No

Have you contributed to test development (i.e., chosen or written test items, contributed to discussion of inclusion of test items, etc.) in at least one nursing course with clinical component within the past academic year?
Yes   No

Do you have access to test item analysis for the classroom tests (i.e., computer generated data related to central tendency, difficulty index, etc.)?
Yes   No

Informed Consent

You are invited to participate in an online survey for a research project conducted ████████████████████████████ IRB requires investigators to provide informed consent to the research participants.

The purpose of this online research study is to examine baccalaureate nursing faculty decision making in classroom test construction, test item analysis, and test revision practices. Specifically the researcher is interested in discovering what are the teacher beliefs, decision making rationales, and decision making processes; and if teacher beliefs, decision making rationales, and decision making processes predict test construction, test item analysis, and test revision practices. You must be at least 18 years old to participate.

If you agree to participate
The survey will take approximately 20 minutes of your time. You will complete a survey about decision making and test development practices. You will not be compensated.

Risks/Benefits/Confidentiality of Data

There are no known risks or discomforts which could cause you to feel uncomfortable, distressed, sad, or tired. There will e no costs for participating. Although your participation in this research may not benefit you personally, it will help us to understand better understanding of current practices and influential factors in classroom test construction, test item analysis, and test revision. You name and email address will not be kept during the data collection phase. This is an anonymous survey. The survey does not ask for any personal information about you. The survey responses will be submitted to the researchers in aggregate without any personal identification. A limited number of research team members will have access to the data during data collection.

Participation or Withdrawal

Your decision to participate or decline participation in this study is voluntary. You may decline to answer any questions and you have the right to withdraw from participation at any time. Withdrawal will not affect your relationship with ████████████ in anyway. If you do not want to participate, click on the "stop survey" arrow or close the browser window.

Contacts

██

Questions about your rights as a research participant

If you have any questions about your rights or are dissatisfied at any time with any part of this study, you can contact, anonymously if you wish, the Institutional Review Board by phone ████████████████████████████████

If you agree to participate in this study please select "yes" If you do not wish to participate select "no".

Thank you in advance for your time and participation!

Yes   No

Demographic and Teaching Background

In this section, the questions are intended to collect information about you, your teaching experience, and the nursing education program you work in. Please indicate the option that best describes you and your nursing education program.

What type of institution do you work in?

| Private liberal arts | Private research intensive | Public | Public research intensive | Proprietary (for profit) institution | Other |

What type of BSN program do you teach in? (Choose all that apply)

| Generic BSN | RN to BSN | Second degree BSN | Accelerated BSN |

What U.S. state is the nursing program in?

How many full time years have you been teaching nursing? (give numerical value)

What is your highest degree completed?

| ADN | BSN | MS in other field | MS in nursing | DNP | EdD | PhD in other field | PhD in nursing |

If you have a master's degree in nursing, what role did your education formally prepare you for? (Choose all that apply)

| Nurse Educator | Nurse Administrator | Nurse Practitioner | Certified Nurse Midwife | Certified Registered Nurse Anesthetist | Clinical Nurse Specialist | Clinical Nurse Leader | None of the above | I do not have a masters in nursing |

Which of the following best describes your position in the BSN nursing program?

| Clinical or non-tenure track | Tenure track- not yet tenured | Tenured | Adjunct or contract faculty |

What is your age in years? (give numerical value)

What is your gender?

| Male | Female |

What is your race/ethnicity? (Choose all that apply)

| Asian | Native Hawaiian | Other Pacific Islander | Black or African American | American Indian or Alaska Native | White | Hispanic Latino | Not Hispanic Latine | More than one race |

Please indicate the amount of course work you have had in classroom test development.

| No course work in test development | Part of one course devoted to test development | One course in test development | More than one course in test development |

Have you ever participated in a professional development program focusing on test development?

Yes   No   Cannot remember

Do you hold certification as a nurse educator (i.e., the CNE credential)?

Yes   No

Introduction Paragraph

The next three sections of the survey relate to test construction, test item analysis, and test revision practices. In answering these questions, please focus on the nursing course with a clinical component in which you contributed to test development within the past year. If you contributed to test development in more than one course with a clinical component, answer the remaining questions about the course for which you had the most involvement.

What was the primary content focus of this course?

Were you the course coordinator of this course?

Yes   No

How many other faculty members contributed to test development in this course? (give numerical value)

Identified Nursing Course

Much of the remaining survey deals with test development practices you have used. When answering these questions please use the course you just described as your frame of reference. The items will refer to this course as the "*identified nursing course*"

Test Construction

The following are components of test construction that faculty can use when constructing a test. Please indicate how often you use each component when developing test items for tests within the *identified nursing course*. Please answer on a scale of 1-7 (1= not at all to 7= all the time).

Please indicate the amount of course work you have had in classroom test development.

No course work in test development	Part of one course devoted to test development	One course in test development	More than one course in test development

Have you ever participated in a professional development program focusing on test development?

Yes	No	Cannot remember

Do you hold certification as a nurse educator (i.e., the CNE credential)?

Yes	No

Introduction Paragraph

The next three sections of the survey relate to test construction, test item analysis, and test revision practices. In answering these questions, please focus on the nursing course with a clinical component in which you contributed to test development within the past year. If you contributed to test development in more than one course with a clinical component, answer the remaining questions about the course for which you had the most involvement.

What was the primary content focus of this course?

Were you the course coordinator of this course?

Yes	No

How many other faculty members contributed to test development in this course? (give numerical value)

Identified Nursing Course

Much of the remaining survey deals with test development practices you have used. When answering these questions please use the course you just described as your frame of reference. The items will refer to this course as the "*identified nursing course*"

Test Construction

The following are components of test construction that faculty can use when constructing a test. Please indicate how often you use each component when developing test items for tests within the *identified nursing course*. Please answer on a scale of 1-7 (1= not at all to 7= all the time).

	1	2	3	4	5	6	7
Course objectives							
Class or unit objectives							
Major content topics							
Specific content topics							
A test blueprint or table of specifications							
The NCLEX-RN test plan							
Peer review of test items							
Higher cognitive levels according to Bloom's taxonomy (i.e., application, analysis, evaluation)							
Clinical context for test items							
Plausible distractors in multiple-choice test items							
Even distribution of correct answer in multiple-choice options							
Use various test item types (i.e., multiple-choice, choose all that apply, fill in the blank, etc.)							

Please indicate on a scale of 1-7 (1= not skilled to 7= very skilled) your skill in test construction.

	1	2	3	4	5	6	7
Overall, how skilled are you at test construction?							

The following items are potential sources of your test items for tests within the identified nursing course. Please indicate the frequency with which you use the sources on a scale of 1-7 (1= never use to 7= always use).

	1	2	3	4	5	6	7
Current course faculty member							
Yourself							
Textbook tests or test item banks							
Review book tests or test item banks							
Previous nursing faculty member-made tests							
Nursing education program-developed test item banks							

Please list any other sources of test items you use that were not identified in the previous item.

Test Item Analysis

The following is a list of different information faculty can obtain about test items after a test is administered. Please indicate how often you use this information after test

administration for tests within the *identified nursing course*. Please answer on a scale of 1-7 (1= not at all to 7= all the time).

	1	2	3	4	5	6	7
The number of students who answered each question incorrectly.							
The number of students who answered each question correctly (difficulty level or p-value).							
A question's ability to discriminate between the high and low scoring students (discrimination index or point biserial coefficient).							
The frequency of distractor choices with each test question.							
The discrimination between the high and low scoring students choosing distractors (distractor discrimination).							
The central tendency (mean, standard deviation) of the student grades on the test.							

Please indicate on a scale of 1-7 (1= not skilled to 7= very skilled) your skill at using test item analyses.

	1	2	3	4	5	6	7
Overall, how skilled are you at using test item analyses?							

Test Revision

The following are actions faculty can perform when revising classroom tests. Please indicate how often you perform these actions during test revision for tests within the *identified nursing course*. Please answer on a scale of 1-7 (1= not at all to 7= all the time).

	1	2	3	4	5	6	7
Use item analysis data when determining to keep or eliminate test questions before finalizing test scores.							
Compare item analysis data for test questions used repeatedly from one term to another.							
Use distractor discrimination to revise test items.							
Use difficulty level of test items to revise test items.							
Assess for linguistic/cultural bias in test items.							
Assess for changes in domain content based upon new research data.							
Assess for outdated language used in test items.							

Change test items to ensure test security.						
Change test items to reflect emphasis in classroom content.						
Change test items to ensure sufficient sampling of content.						

Please indicate on a scale of 1-7 (1= not skilled to 7= very skilled) your skill at test revision.

	1	2	3	4	5	6	7
Overall, how skilled are you at test revision?							

Teacher Beliefs

The following statements pertain to your beliefs related to evaluation of students in the classroom. Please indicate on a scale of 1-7 (1= least descriptive to 7= most descriptive) the extent to which these statements describe your beliefs.

	1	2	3	4	5	6	7
What is best for the students is the major consideration.							
In this nursing education program, nursing faculty members are expected to follow their own personal and moral beliefs during student evaluation.							
During student evaluation, nursing faculty members look out for the student's good.							
Each nursing faculty member in this nursing education program decides for himself or herself what is right or wrong in student evaluations.							
It is important to follow the nursing education program's evaluation rules and procedures during student evaluation.							
The most important concern in this nursing education program during student evaluation is the individual faculty member's sense of right and wrong.							
Successful student evaluation occurs by strictly following the nursing education program's evaluation rules and procedures.							
In this nursing education program, nursing faculty members are guided by their own personal ethics.							

During student evaluation, it is expected that nursing faculty members will always do what is right for the student.								
Nursing faculty members in this nursing education program adhere to the program's rules and procedures of student evaluation.								
The most important concern in student evaluation is the good of the student body as a whole.								
In student evaluations, the best practices of student evaluation are the major consideration.								
Nursing faculty members in this nursing education program are expected to comply with best practices of student evaluation over and above other considerations.								
In this nursing education program, the most efficient way to move students through the nursing program is the right way for student evaluations.								
In this nursing education program, nursing faculty members are expected to strictly follow best practices of student evaluations.								
In this nursing education program, nursing faculty members are expected to above all move student efficiently through the nursing program.								
In this nursing education program, the first consideration is whether a decision regarding student evaluation violates any best practices of student evaluation.								
The major concern of nursing faculty members in this nursing education program is always what is best for the students.								
All nursing faculty members are expected to stick by the nursing education program's evaluation rules and procedures.								

You are almost done with the survey. Thank you for your perseverance.

Decision Making

In this section, the items pertain to factors that may influence your decision making about test development. Please indicate how important each of these factors are in

influencing your decisions regarding test construction, test item analysis, and test revision.

	Very important	Moderately important	Minimally important
Tradition and past practice			
Research about the effectiveness of various strategies			
Educational soundness or congruence with educational standards			
Peer pressure			
Regulatory standards			
NCLEX-RN pass rates			
Time needed to develop test			
Time needed to implement test			
Time needed to conduct test item analysis			
Time needed to revise test items			
Material and equipment needed for test construction			
Openness to creativity			
Incentives (or lack thereof) for innovation			
Legal defensibility			
Personal ethics			

Decision Making Processes

People differ in the way they go about making decisions. Please indicate how you make decisions by selecting for each item the response which best fits your usual style.

When making decisions -

	True for me	Sometimes true	Not true for me
I feel as if I'm under tremendous time pressure when making decisions.			
I like to consider all of the alternatives.			
I prefer to leave decisions to others.			
I try to find out the disadvantages of all alternatives.			
I waste a lot of time on trivial matters before getting to the final decisions.			
I consider how best to carry out the decision.			

Even after I have made a decision I delay acting upon it.			
When after making decisions I like to collect lots of information.			
I avoid making decisions.			
When I have to make a decision I wait a long time before starting to think about it.			
I do not like to take responsibility for making decisions.			
I try to be clear about my objectives before choosing.			
The possibility that small things might go wrong causes me to swing abruptly in my preferences.			
If a decision can be made by me or another person I let the other person make it.			
Whenever I face a difficult decision I feel pessimistic about finding a good solution.			
I take a lot of care before choosing.			
I do not make decisions unless I really have to.			
I delay making decisions until it is too late.			
I prefer that people who are better informed decide for me.			
After a decision is made I spent a lot of time convincing myself it was correct.			
I put off making decisions.			
I cannot think straight if I have to make a decision in a hurry.			

Classroom Realities

In this section, the questions are intended to collect information regarding your classroom environment and students.

Please select all the characteristics of the majority of students enrolled in the *identified* nursing course. (Choose all that apply)

- Prior preparation in nursing (e.g., PN, RN)
- Prior preparation in health care (e.g., paramedic, medical corpsman)
- Racially diverse
- Traditional age (i.e., 18-22 years old)
- Adult learners
- English Language Learners

Female

Equal mixture of females and males

What are the average number of students enrolled in the *identified nursing course* per term? (give numerical value)

Please indicate the primary method of class delivery for the *identified nursing course*.

Face-to-face   On-line   Hybrid (web enhanced face-to-face)

What is the method of test delivery for classroom tests in the *identified nursing course*?

Paper and pencil   Computer   Both

What is the setting for testing for the *identified nursing course*? (Choose all that apply)

Computer lab   Classroom with   Classroom with   Testing Center
                    laptops   paper and pencil

What, if any, technologies are used in test delivery for the *identified nursing course*? (Choose all that apply)

Computers   Learning   Lockdown browsers   Other
                    Management
                    System

External Factors

In this section, the questions are intended to collect information regarding individuals or agencies apart from you who influence student assessment and evaluation.

What is the most recent NCLEX-RN pass rate for your program? (give numerical value)

In the past 5 years, have there been any concerns expressed to faculty from administration about the NCLEX-RN pass rates for your nursing education program?

Yes   No   Do not know

What is the nursing accrediting agency used at your nursing education program?

CCNE   NLNAC   None   Do not know

What is the mandated NCLEX-RN pass rate by your local state board of nursing? (give numerical value)

The following items pertain to who or what most often determines student assessment and evaluation strategies within your nursing program. Please indicate on a scale of 1-7.

114

(1= not influential to 7= very influential) who or what is most influential in determining student assessment and evaluation strategies.

	1	2	3	4	5	6	7
Individual faculty member							
Course faculty team							
Committee in nursing program							
Institution (e.g., university academic policy committee)							
Accrediting bodies							
State board of nursing							