All Electronic Theses and Dissertations

2021

Use of Best Practices in Exam Item Creation, Analysis, and Revision: Nursing Faculty's Knowledge, Use, and Implementation Barriers and Facilitators

Julie Anne De Haan
Bethel University

Follow this and additional works at: https://spark.bethel.edu/etd

Use of Best Practices in Exam Item Creation, Analysis, and Revision:

Nursing Faculty's Knowledge, Use, and Implementation Barriers and Facilitators

By:

Julie Anne Baas De Haan

A dissertation submitted to the faculty of Bethel University

in partial fulfillment of the requirements for the degree of

Doctor of Education

St Paul, MN

2021

Dissertation Advisor: Joel Frederickson

Readers: Connie Clark

Diane Dahl

Abstract

The ultimate goal of nursing education is to prepare graduates to provide excellent, safe patient care. The implementation of exam item best practices can significantly contribute to this process. This study examined nursing faculty's use of best practices for exam item creation, analysis, and revision and best practice barriers and facilitators. Through a quantitative, cross-sectional, descriptive correlational research study, the research demonstrated that faculty inconsistently use exam item best practices. Faculty research participants identified ways to enhance the use of best practices and suggestions to nursing leadership were developed based on the data. With faculty development and an outlined support system in place, faculty will be more equipped to implement exam item best practices and contribute to the goal of well-prepared graduates.

Dedication

For Joel
Two are better than one,
because they have a good return for their labor:
if either of them falls down,
one can help the other up…
A cord of three strands is not quickly broken.
Ecclesiastes 4:9-10a, 12b (ESV)

Acknowledgments

My friend Amy compared writing a dissertation to birthing an elephant. I birthed this dissertation surrounded and supported by many people who deserve recognition.

- My husband Joel, who encouraged me when I wanted to quit and picked up more than his share of the household responsibilities.

- My parents, Julian and Norma Baas, who modeled what it means to be a Christ follower and lifelong learners

- My four children, Marlynda, Mitchell, Toben, and Tracelyn, who acted as cheerleaders and graciously gave me space to do my work

- My dissertation support group, Allison, Claudia, and Peggy, as dysfunctional as we may have been at times, you kept me going, you got it.

- My dissertation advisor Dr. Joel Frederickson, who saved me from drowning in data and offered timely advice and encouragement

- My dissertation committee, Diane Dahl and Connie Clark, who gave great feedback and challenged my thinking

- My colleagues, who listened and inspired

- My God, who is my peace, my shelter, my purpose for living. Solo Deo Gloria

Table of Contents

List of tables

List of Figures

Chapter 1: Introduction

Throughout nursing programs, faculty tailor multiple-choice questions (MCQs) to assess students' learning and to prepare graduates for taking the National Council Licensure Examination (NCLEX) to obtain licensure. Successful completion of the NCLEX exam serves as a gateway to practice for an entry-level nurse. Nursing students' exams need to effectively assess students' ability to apply course content to nursing practice and provide students with practice answering higher-level thinking questions, such as those on the NCLEX (Birkhead, Kelman, Zittel, & Jatulis, 2018; Cox, 2019; Ibrahim, 2019; Mager, Beauvais, & Kazer, 2017; Tarrant & Ware, 2012). Furthermore, students' success on exams is tied to students' progression through the nursing program (Birkhead et al., 2018; Cox, 2019; Hijji & Mahmoud, 2017; Omer, Abdulrahim, & Albalawi, 2016). With academic and career success contingent upon students' performance on exams, nursing educators have a professional and ethical responsibility to construct exam items that align with best practices (Libner & Kubala, 2017; National League for Nursing, 2012; O'Rae, Hnatyshyn, Bock, Mannion, & Patek, 2019; Tarrant & Ware, 2012). As Boud (1995) stated, "Students can, with difficulty, escape from the effects of poor teaching, they cannot (by definition if they want to graduate) escape the effects of poor assessment" (p. 35).

Although nursing educators employ several types of assessment, faculty utilize MCQs more consistently than other means of assessment (Birkhead et al., 2018). Students' performance on exams determines their progression in the program and also provides a means for students to practice taking exams before taking the NCLEX to obtain licensure. The NCLEX is written to determine if graduates have the baseline thinking skills needed for safe patient care in entry level practice in nursing. Nurses must exercise clinical judgment—a high level of thinking—when caring for patients. Putting this within the context of Bloom's taxonomy

(Bloom, 1956), this type of thinking requires higher-order thinking practices such as evaluation and analysis in order to create and implement a plan of care for their patients' individualized situations. The NCLEX exam items exclusively test higher-order thinking. Students' proficiency at answering higher-order thinking multiple-choice questions increases when nursing faculty provide more opportunities for students to interact with these types of questions. The implementation of exam item best practices would increase the quality of exam items used in nursing education and better prepare students for taking the board exam and for entry-level practice.

The majority of nursing exams include multiple-choice questions (MCQ) (Birkhead et al., 2018; Killingworth et al., 2015). Killingsworth et al. (2015) explored what best practices for exam construction nursing faculty chose to use. The best practices in MCQ construction were also discussed by Cox (2019) in an integrative review of 28 articles written between 2008 and 2017 regarding the use of multiple-choice testing in nursing programs. Along with following item writing guidelines, Cox (2019) advocated for posttest item analysis and review of items for flaws. Flawed items potentially cause students to fail exams that they could have passed (Cox, 2018). Hijji (2017) also examined MCQs for item-writing flaws and found that 91.8% of the items contained one or more flaws (p. 492). Omer et al. (2016) supported Cox's (2019) findings as well. Flaws in multiple-choice writing can make items easier or harder. Flawed items also increase the possibility of a construct-irrelevant variance, which affects students' performance on the exam as a whole. Omer et al. (2016) found that students who were on the lower academic achievement border passed the exam at a higher rate if flawed items were included while higher-achieving students' performances were lower on the exam. A higher percentage of flawed items

significantly jeopardizes assessment of learning and places into question the score earned by the student.

Libner and Kubala (2017) emphasized the need for faculty to improve exam construction skills when the NCLEX first-time pass rates of nursing schools dipped below 75% in Illinois and in numerous nursing programs across the nation after changes made to the NCLEX in 2014. Better exam construction skills were identified as one of several steps these institutions were required to implement as part of the Illinois Board of Nursing's remediation plan. Mager et al. (2017) expanded the discussion of the Illinois Board of Nursing ten-step plan and detailed how they improved the exam construction skills at their institution when NCLEX pass rates fell.

Many nursing educators lack formal preparation regarding the creation, analysis, and revision of exam items which results in flawed items, poor indices of difficulty (and poor reliability), and questionable differentiation indices (leading to poor validity) (Birkhead et al., 2018; Cox, 2019; D' La & Visbal-Dionaldo, 2017; Hijji & Mahmoud, 2017; Killingsworth et al., 2015; Mager et al., 2017; O'Rae et al., 2019; Obon & Rey, 2019). Thus, a higher number of flawed items on an exam directly and negatively affects the overall exam statistics for validity and reliability. Nursing faculty need training on how to evaluate exam items for validity and reliability in order to revise items effectively. Better individual item statistics can increase overall exam validity and reliability.

The NCLEX test bank exclusively consists of higher-level cognitive items which test students' ability to apply, evaluate, and analyze. Items that assess students' understanding and comprehension, lower-level cognitive processes, are not included. Therefore, the student's ability to answer higher-level cognitive items should be scaffolded throughout the nursing program to allow students to develop their ability to take exams with higher-level thinking items.

14

Many nursing programs use blueprints to ensure that exam items are written at higher cognitive levels and are statistically valid and reliable items (Ibrahim, 2019; Oermann & Gaberson, 2021; Tarrant & Ware, 2012). A robust testing policy holds faculty accountable for editing items and creating exams that remain statistically valid and reliable and requires exam blueprints that address the consistent use of higher-level thinking items (Oermann & Gaberson, 2021; Schroeder, 2013). Testing policies also increase exam quality consistency across the curriculum (Barton et al., 2014).

Problem Statement

Educators from several disciplines, including nursing, have identified best practices for exam creation (Coffman et al., 2010; Cox, 2019; D' La & Visbal-Dionaldo, 2017; Halstead, 2013; Ibrahim, 2019; Killingsworth et al., 2015; Lavin & Rosario-Sim, 2013; Libner & Kubala, 2017; Mambwe, 2017; Naeem et al., 2012; O'Rae et al., 2019; Obon & Rey, 2019; Oermann & Gaberson, 2021; Ray et al., 2018; Rudolph et al., 2019; Schroeder, 2013; Tarrant & Ware, 2012; White & Heitzler, 2018). Faculty use of best practices for exam creation remains inconsistent (Clifton & Schriner, 2010; Hijji & Mahmoud, 2017; Killingsworth et al., 2015; O'Rae et al., 2019). There are several barriers that impede nursing educators from using exam item best practices, such as a knowledge gap, lack of faculty development, time constraints, budget constraints, lack of policies to support best practices, and insufficient administrative and administration support (Birkhead et al., 2018; Halstead, 2013; Ibrahim, 2019; Lavin & Rosario-Sim, 2013; Obon & Rey, 2019; Ray et al., 2018; Tarrant & Ware, 2012). Conversely, there are also several potential support systems to help nursing faculty utilize exam item best practices, including faculty development, the use of exam blueprints, clear detailed policies, department item developer position to support faculty, peer feedback mechanisms, mentoring for new

faculty, and administrative and administration (leader) support (Birkhead et al., 2018; Halstead, 2013; Ibrahim, 2019; Lavin & Rosario-Sim, 2013; Obon & Rey, 2019; Ray et al., 2018; Tarrant & Ware, 2012).

Nurse educators can be overwhelmed by the prospect of implementing best practices for item creation, analysis, and revision. Although several researchers and assessment experts from higher education have suggested activities that best guide the creation, analysis, and revision of exam items (Ibrahim, 2019; Killingsworth et al., 2015; Naeem et al., 2012; National League for Nursing, 2012; Oermann & Gaberson, 2021; O'Rae et al., 2019; Rudolph, et al., 2019), one defined list of best practices have not been adopted by nursing education organizations. Without such a recommendation, nursing faculty lack specific guidance regarding the implementation of best practices for item creation, analysis, and revision. Furthermore, few research studies examined what exam item best practices faculty consistently use.   Ibrahim (2019), Lavin and Rosario-Sim (2013), and Naeem et al.'s (2012) research demonstrated an increase in exam quality when faculty utilized best practices for exam creation. Killingsworth et al. (2015) and O'Rae et al. (2019) explored the process by which faculty create exams and touched upon which best practices faculty employ, but did not correlate these findings to the context in which faculty practice. Faculty context includes the demographic, individual, environmental, and social faculty characteristics. Therefore, the faculty context or faculty factors that facilitate or prevent best practices for exam creation remain unknown.

Purpose Statement

The first purpose of this study was to examine relationships between nursing faculty's demographic, individual, environmental, and social factors and their knowledge and use of exam

item best practices. The second purpose was to examine what nursing faculty identified as potential barriers and facilitators of exam item best practices implementation.

Research Questions

RQ1: What relationship, if any, exists between faculty's demographic factors (i.e., age, gender) and their score on the exam item best practices list?

H_o1: There is no relationship between faculty's demographic factors and their utilization of exam item best practices score.

RQ2: What relationship, if any, exists between faculty's individual factors (i.e., level of education, length of time since highest degree completion, years of teaching experience, frequency of writing exam items, certification as Certified Nurse Educators (CNE)) and their score on the exam item best practices list?

H_o2: There is no relationship between faculty's individual factors and their utilization of exam item best practices score.

RQ3: What relationship, if any, exists between faculty's environmental factors (i.e., type of institution at which the participant works, department has a testing policy, administrative support available, administration support of exam creation best practices, and availability of funds for faculty development) and their score on the exam item best practices list?

H_o3: There is no relationship between faculty's environmental factors and their utilization of exam item best practices score.

RQ4: What relationship, if any, exists between faculty's social factors (i.e., formal and/or informal mentor) and their score on the exam item best practices list?

H_o4: There is no relationship between faculty's social factors and their utilization of exam item best practices score.

RQ5: What is the frequency and variation of ratings of the faculty-identified barriers and facilitators of exam item best practice implementation?

RQ6: What associations exist between faculty-identified barriers and facilitators of exam item best practice implementation and faculty's demographic, individual, environmental, and social factors?

Significance of Study

Nursing faculty are content experts. As nurses, they know how to care for patients, how to promote health, and how to be client advocates; however, nurses are not necessarily prepared to teach nursing. Nursing educators can obtain a variety of graduate degrees to qualify for a faculty position. Nurse practitioners' degrees usually focus on a specific population, such as geriatrics, pediatrics, family practice, psychiatric, and other areas. Nurse practitioners expertly care for their patient population. Other faculty may obtain advanced degrees in nursing leadership or have an advanced degree related to nursing research. Some faculty acquire a post-graduate degree in nursing education. Although nursing education programs address general assessment practices and assessments specific to clinical, faculty may or may not receive training in exam item construction, analysis, and revision. National nursing organizations have not adopted a set of best practices for exam item development, but rather have addressed fairness in testing issues that apply to the various methods of assessment across the nursing curriculum (National League for Nursing, 2012). Many exam items in nursing and other health sciences have poor item statistics and contain numerous flaws (Birkhead et al., 2018; D' La & Visbal-Dionaldo, 2017; Mambwe, 2017; Naeem et al., 2012; Obon & Rey, 2019; Omer et al., 2016; Ray et al., 2018; Rudolph et al., 2019). Therefore, faculty need to ask whether student assessments are valid. Additionally, with the emphasis on higher-order thinking and clinical judgment

development, nursing educators need to understand whether they are preparing their students well for practice.

Faculty administer assessments to gauge student progression and achievement of course objectives. Nurse educators also use assessment to develop and measure higher-order thinking. Well-designed MCQs require the student to grow in and use higher-order thinking processes. Nurses practice higher-order thinking when caring for patients. In nursing, the use of higher-order thinking processes is referred to as clinical judgment. Nursing education trains graduates to practice clinical judgment in order to provide the best care for patients and ensure patient safety. Clinical judgment remains central for safe, effective patient care (Betts et al., 2019; Dickison et al., 2019). Although nursing faculty utilize a variety of assessments, MCQ continues to be the most commonly chosen assessment method (Birkhead et al., 2018). The NCLEX exclusively utilizes higher order thinking items with the majority of the items being multiple choice or multiple response questions. Multiple response questions contain more than one correct answer and students select all the potential answers that apply to the stem (Oermann & Gaberson, 2021). Things that undermine the quality of exams include low reliability and validity statistics, ineffective distractors, and item flaws. Faculty use of exam item best practices promotes and sustains high-quality exam items. However, little research has focused on which best practices faculty use (Cox, 2019; Killingworth et al., 2015; O'Rae et al., 2019, Rudolph et al., 2019).

As mentioned previously, researchers from different health sciences identified exam item best practices and the literature repeatedly points to the need for best practice implementation (Birkhead et al., 2018; D' La & Visbal-Dionaldo, 2017; Mambwe, 2017; Naeem et al., 2012; Obon & Rey, 2019; Omer et al., 2016; Ray et al., 2018; Rudolph et al., 2019). Some research

has been done regarding what pieces of best practices faculty use (Killingsworth et al., 2015; O'Rae et al., 2019). However, a gap in the literature exists regarding what faculty identify as barriers and facilitators of exam item best practice implementation (Cox, 2019; Killingworth et al., 2015; O'Rae et al., 2019, Rudolph et al., 2019).

Through this study, data were gathered regarding the exam item best practices faculty implement regularly and examined whether there were relationships between faculty factors and use of exam item best practices. These data provide nursing department leaders with a look at what best practices are being regularly used by the participants and what best practices are not being implemented. Nursing department leaders may employ the best practices list compiled for this study to evaluate what their faculty use. Additionally, nursing departments may become aware of the need for further faculty development regarding exam item creation (Ibrahim, 2019; Naeem et al., 2012; Oermann & Gaberson, 2021; Tarrant & Ware, 2012).

Individual faculty may have self-evaluated their exam item creation practices as they participated in the survey. Self-reflection can begin the journey to change. Furthermore, awareness of best practices may provide the impetus to implement unutilized processes or seek out educational opportunities (Boud & Soler, 2015).

In addition to indicating which exam item creation best practices faculty regularly utilize, participants identified which facilitators and barriers for the implementation of exam item best practices they consider most prevalent in their setting. These data could prove instrumental in the development of nursing departments' policies regarding exam item best practices, the use of blueprints, establishing peer feedback mechanisms, and other facilitators of best practices (Birkhead et al., 2018; Halstead, 2013; Ibrahim, 2019; Lavin & Rosario-Sim, 2013; Obon & Rey, 2019; Ray et al., 2018; Tarrant & Ware, 2012). Financial and time constraints inhibit the

implementation of all possible facilitators (Cox, 2019; Ibrahim, 2019; Killingsworth et al., 2015; O'Rae et al., 2019).   Knowing what faculty identify as barriers and facilitators, nursing department leaders can make better-informed decisions regarding the use of resources to improve exam item quality. Ultimately, the implementation of better exam items improves the preparation of graduates, which leads to better, safer, more efficient patient care (Oermann & Gaberson, 2021).

Definition of Terms

   Exam blueprint.   An exam blueprint consists of faculty-identified course and module objectives to be tested, the distribution of item type, cognitive level, the associated step of the nursing process, and client need categories (Oermann & Gaberson, 2021; Schroeder, 2013). The creation of the blueprint prior to the development of the items ensures that the assessment will measure the important concepts in a manner that utilizes higher-order thinking.

   Testing policy.   In order to promote consistency across the curriculum, nursing departments should adopt a testing policy. A robust testing policy addresses both internal and external evaluation topics. Internal topics include the use of blueprints and best practices for item creation, benchmarks for item analysis, the environment in which students will test, guidelines for test review with students, address exam security, and provide direction on how to attend to various student circumstances such as tardiness, illness, and cheating (Oerman & Gaberson, 2021; Schroeder, 2013; Tarrent & Ware, 2012).

   Item difficulty (DIF).   Item difficulty (DIF) refers to the proportion of students who correctly answer an item. This is also described as item reliability. Education experts regard DIF between 20%-90% as good and acceptable, while DIF between 40%-60% is considered excellent. The higher the DIF, the easier the item. Therefore, items with a DIF above 90% or

below 20% should be evaluated for revision or elimination (D' La & Visbal-Dionaldo, 2017; Obon & Rey, 2019).

Discrimination index (DI). A discrimination index (DI) is assigned to each item posttest and is based upon the percentage of students who collectively scored in the upper quartile on the exam compared to those who scored in the lower quartile. The DI assists educators to determine the effectiveness or validity of an item in multiple-choice exams. This statistic is reported from -1.0 to 1.0. Items with a 1.0 result indicate perfect discrimination between high and low-achieving students with all the high-scoring students getting the item correct and all the low-scoring students getting the item wrong. A negative result indicates that more students who scored lower on the exam got this item correct than those who achieved a higher score. Items with a low or negative DI need to be evaluated for flaws or distractor efficacy (D' La & Visbal-Dionaldo, 2017; Obon & Rey, 2019).

Distractor efficiency (DE). Distractor efficiency is based upon the number of times a distractor was chosen. The careful design of distractors, or the wrong answer, contributes to the quality of the MCQ. If distractors are not chosen by any students, educators need to evaluate the plausibility of the distractors to promote higher levels of thinking and ensure that the distractor does not contribute to construct-irrelevancy (D' La & Visbal-Dionaldo, 2017; Obon & Rey, 2019; Ray et al., 2018).

Exam reliability. Exam reliability is determined by evaluating the internal consistency of the items. The Kruger- Richardson formula (as known as the KR20) is a coefficient that indicates internal consistency, or homogeneity, on a 0-1.0 scale. The formula looks at the consistency of the items DIF, the number of items on the exam, and the standard deviation. The closer that the KR20 is to 1.0, the higher the internal consistency of the exam. In general, a

KR20 result of >0.8 is considered excellent (Obon & Rey, 2019; Rudolph et al., 2019; Tarrant & Ware, 2012).

Overview of the Remainder of the Study

The remainder of this study is divided into four chapters. Chapter 2 reviews the literature related to the use of multiple-choice questions in nursing education and exam item best practices identified in nursing education publications, as well as other health sciences. The potential barriers and facilitators of exam item best practices identified in the literature are discussed at the end of the chapter. Chapter 3 outlines the methodology for survey creation, data collection, and analysis and discusses the conceptual framework for this study. The results of this study are presented in Chapter 4. Chapter 5 includes a discussion and implications of the results with recommendations for further research. Appendices and references are included at the end of this dissertation.

Chapter 2: Literature Review

Introduction

Throughout nursing programs, multiple-choice questions (MCQs) are tailored to assess students' learning and to prepare graduates for taking the National Council Licensure Examination (NCLEX) to obtain licensure. Nursing student exams need to effectively assess students' ability to apply course content to nursing practice and provide students with practice taking higher-level thinking questions such as those on the NCLEX (Birkhead et al., 2018; Cox, 2019; Ibrahim, 2019; Mager et al., 2017; Tarrant & Ware, 2012). Furthermore, student success on exams is tied to student progression (Birkhead et al., 2018; Cox, 2019; Hijji & Mahmoud, 2017; Omer et al., 2016). With academic and career success contingent upon exams, nursing educators have a professional and ethical responsibility to construct reliable and valid items in order to create exams that are internally consistent and repeatable (Libner & Kubala, 2017; O'Rae et al., 2019; Tarrant & Ware, 2012).

Many nursing educators lack formal preparation for this aspect of student assessment which results in flawed items, poor indices of difficulty (poor reliability), and questionable differentiation indices (leading to poor validity) (Birkhead et al., 2018; Cox, 2019; D' La & Visbal-Dionaldo, 2017; Hijji & Mahmoud, 2017; Killingsworth et al., 2015; Mager et al., 2017; O'Rae et al., 2019; Obon & Rey, 2019). This chapter will provide a summary of some of the literature regarding the use of MCQs in nursing education, and explore the literature for the best practices regarding pre-exam framework, the use of blueprints for exam construction, the development of item stems and distractors, the evaluation of MCQ validity and reliability, exam reliability, and the elements of exam review prior to administration. The discussion regarding blueprints will include the importance of assessing students' higher-level cognition by creating

items that reflect the analyzing and applying cognitive levels of Bloom's taxonomy (Ibrahim, 2019; Killingsworth et al., 2015; Mager et al., 2017; Mambwe, 2017; Tarrant & Ware, 2012). The NCLEX test bank consists of higher-level cognitive items. Items that test understanding and comprehending lower-level cognitive processes are not included on the NCLEX and students should be practicing higher order thinking throughout the nursing curriculum (Oermann & Gaberson, 2021).

Ethical Mandate

The National League for Nursing (2012) issued *Fair Testing Guidelines for Nursing Education.* The first guideline states:

> Faculty have an ethical obligation that both tests and the decisions based on tests are valid, supported by solid evidence consistent across their programs, and fair to all test takers regardless of age, gender, disability, race, ethnicity, national origin, religion, sexual orientation, linguistic background, testing style, and ability, or other personal characteristics. (p. 1)

The remainder of the document further outlines how to select appropriate tests, inform test-takers, administer and score tests, as well as how to report and interpret test results. The document concludes with recommendations regarding establishing a fair testing environment (National League for Nursing, 2012). The guidelines clearly support the use of best practices when creating and evaluating exam items and equate their use with the ethical obligation nursing faculty hold.

Use of Multiple-Choice Questions (MCQ) in Nursing Exams

Although nursing education utilizes a wide variety of assessments, exams constructed with MCQs remain prevalent. Killingsworth et al. (2015) stated that MCQ exams formed the

most influential form of assessment used by nursing educators which determined course grades and program progression. Birkhead et al. (2018) examined nursing educators' use of MCQ by sending out an online survey to nursing educators. They enquired about the use of MCQs in their programs and how achievement on multiple-choice exams impacted the student's grade in the course. The majority of respondents (74%) indicated that at least 80% of exam items were MCQs on a typical exam. Sixty-five percent of the respondents indicated that scores from multiple-choice exams accounted for at least 80% of the course grade. Birkhead et al. (2018) confirmed Killingworth et al.'s (2015) findings that MCQ exams were the most influential assessment technique used by nursing faculty.

Impact on Student Grades, Progression, and NCLEX Success

Killingsworth et al. (2019) focused on which best practices for exam construction nursing faculty chose to use. Their results indicated that their participants ($n = 127$) frequently used 22 of the 26 best practices. Best practices in MCQ construction were also discussed by Cox (2019) in her integrative review of 28 articles written between 2008 and 2017 regarding the use of multiple-choice testing in nursing programs. Cox (2019) advocated for posttest item analysis and review of items for flaws. Posttest item analysis will be discussed in a separate section. Flawed items potentially cause students to fail exams that they could have passed (Cox, 2019). Hijji (2017) examined MCQs for item-writing flaws and found that 91.8% of the items contained one or more flaws. This high percentage of flawed items significantly jeopardizes assessment of learning and places into question the score earned, or not earned, by the student. Omer et al. (2016) supported Cox's (2019) statements. Omer et al. (2016) discovered that flaws in multiple-choice questions can make items easier or harder. Flawed items also increase the possibility of construct-irrelevant variance which affects students' performance on the exam as a whole.

Construct-irrelevant items include factors that are not relevant to the concept being tested. In their study, students who were on the lower academic achievement border passed the exam at a higher rate if flawed items were included; while higher-achieving students' performances were lower on the exam.

Libner and Kubala (2017) emphasized the need for faculty to improve exam construction skills when the NCLEX first-time pass rates of nursing schools dipped below 75% in Illinois in 2014. Better exam construction skills were identified as one of several steps these institutions were required to implement as part of the Illinois Board of Nursing's (BON) remediation plan. Mager et al. (2017) expanded the discussion of the Illinois BON's ten-step plan and detailed how they improved the exam construction skills at their institution when NCLEX pass rates fell. They discussed test item analysis which increased their awareness of MCQ validity and reliability and improved their MCQ item construction skills. After implementing the ten-step plan, the first-time pass rate increased by 10% at Mager et al.'s (2017) institution (p. 283).

MCQ Validity and Reliability

Numerous resources identified the need for training regarding the construction of MCQs, how to review items for flaws, and how to analyze item statistics (Birkhead et al., 2018; Cox, 2019; D' La & Visbal-Dionaldo, 2017; Ibrahim, 2019; Libner & Kubala, 2017; Obon & Rey, 2019; Tarrant & Ware, 2012). The importance of posttest statistical analysis included discussion of item difficulty, discrimination index, and distractor efficiency.

With the item and exam analysis data in hand, nursing faculty can make decisions regarding the reliability and validity of items and exams as a whole. Obon and Rey (2019) completed a descriptive study of 194 MCQs in one nursing course in which they examined item difficulty, discrimination index, distractor efficiency, and exam reliability by KR20. They found

that 59% of items required revision based on item difficulty, 63% of items required revision based on discrimination index, and 30% of the distractors emerged as non-functioning and therefore should be revised. Tarrant and Ware (2012) emphasized the importance of item analysis during posttest reviews. They included difficulty index, discrimination index, distractor frequency, distractor discrimination, and test reliability (KR20) in the statistical data of the posttest review. Tarrant and Ware (2012) stated, "All test items need editing and refinement" (p. 103).

Rigorous MCQs ensure that students are assessed on the intended content. White and Heitzler (2018) developed rigorous MCQs as part of their research to decrease grade inflation, which links to Ray et al.'s (2018) discussion regarding construct representation and construct-irrelevant variance. White and Heitzler (2018) defined rigorous MCQs as those that followed best practices for MCQ construction and were regularly reviewed posttest so that both high quality and poor performing items could be identified, revised, or eliminated.

Exam Item Best Practices

Testing Policy. Oermann and Gaberson (2021) stressed the importance of a detailed testing policy implemented across the curriculum. A comprehension testing policy provides a framework that guides faculty in exam creation and a consistent administration of assessments for students A testing policy should address general exam construction with guidelines for types of items, number of items per exam, time allotment for the exam, exam administration, and exam security. Furthermore, the policy should outline best practices for exam item writing, blueprinting, exam analysis, and guidelines for sharing exam results with students (Oermann & Gaberson, 2021; Schroeder, 2013).

Pre-exam Planning. Prior to the assessment, faculty decide what type of assessment would best evaluate students at a given point in the curriculum, both within the course and program. The assessment may be formative or summative, taken individually, or collaboratively, or both. Faculty consider the timing of the assessment, not only when in the course the assessment will take place, but also the amount of time allotted. Decisions regarding *where* the exam will take place also must be made. Will the exam be given in a classroom, online, in the nursing laboratory, or be a take-home exam? Additionally, faculty consider what opportunities students will have to receive feedback. The pre-planning establishes the context from which the assessment will be created (Bearman et al., 2014; Boud & Soler, 2015; Cox, 2019; Oermann & Gaberson, 2021)

Exam Blueprints. An exam blueprint consists of faculty-identified course and module objectives to be tested, the distribution of item type, cognitive level, nursing process, and client need categories (Eweda et al., 2020; Oermann & Gaberson, 2021; Schroeder, 2013). The creation of the blueprint prior to the development of the items ensures that the assessment will measure the important concepts in a manner that utilizes higher-order thinking.

Blueprint construction. To help facilitate high-quality exam construction, many researchers and assessment experts encourage the use of exam blueprints (Cox, 2019; Eweda et al., 2020; Killingsworth et al., 2015; Mambwe, 2017; Oermann & Gaberman, 2021; O'Rae et al., 2019; Ray et al., 2018). Oermann and Gaberman (2021) compared the use of exam blueprints to those used for home construction and stressed the importance of developing a specific framework for an exam prior to creating the exam itself. The blueprint should include the main topics or course objectives that will be assessed, the level of thinking skills needed to complete the exam, and the weight of each topic in proportion to course content (Oermann &

29

Gaberman, 2021, Ray et al., 2018). Creating a blueprint that aligns with the students' level and further assesses clinical judgment formation should be the first step in the creation of an exam (Eweda et al., 2020).

   Blueprints and improving exam validity. Beyond ensuring that exams align with learning objectives and clinical judgment development, Ray et al. (2018) supported the use of blueprints as a means to ensure maximum exam validity by reducing two validity threats: construct representation and construct-irrelevant variance. O'Rae et al. (2019) identified four needs of nursing faculty in order to construct high-quality exams. One of these needs stressed the development of an assessment blueprint that reflects expected student development across the curriculum. This correlates with Ray et al.'s (2018) findings of decreasing exam validity threats by using blueprints that align with specific course content and student knowledge development as they progress through nursing programs. Many researchers also purport the use of blueprints to ensure that exam items are written at higher cognitive levels (Ibrahim, 2019; O'Rae et al., 2019; Tarrant & Ware, 2012).

   Inclusion of Bloom's taxonomy. Mambwe (2017) discussed the use of blueprints to ensure that exam construction contained higher-level items in her research which looked at nursing educator knowledge of Bloom's taxonomy when constructing exams. The literature supports the use of higher-level exam items to develop critical thinking skills (Scully, 2017). If nursing educators do not construct exam questions that test students at the higher levels of Bloom's taxonomy, evaluation of higher-level thinking skills cannot occur. Without regular scaffolding of formative assessment, the danger of graduating nurses without crucial critical thinking skills increases (Scully, 2017). Tarrant and Ware (2012) also stressed that test blueprints help ensure that an exam is valid and reliable by outlining the proportion of questions

30

across content areas and indicating the cognitive level at which the items should be written. The assertions of Ray et al. (2018) added to this conversation when discussing construct-irrelevant variance by eliminating flawed item formats such as an inappropriate level of difficulty, which includes items that are too easy or too difficult. The authors emphasized that item difficulty level should be selected based upon the learner level and the learning objectives being assessed.

Variety of item difficulty. Numerous experts and researchers stressed the importance of choosing items with a range of difficulty when designing a course exam (Cox, 2019; Ibrahim, 2019; Omer et al., 2016; Oermann & Gaberson, 2021; Ray et al., 2018). A range of difficulty allows for completion of the exam within the scheduled time frame (Ray et al., 2018) and for differentiation of student performance, otherwise known as norm referencing (Oermann & Gaberson, 2021). If the items are too easy or too difficult, the exam will not discriminate between students' knowledge and ability. Faculty may design exams to determine minimal competency and set a benchmark score to determine progress. The criterion-referenced exams should consist of moderate to high difficulty questions. Criterion-referenced exams are often administered at the end of a course or program. The NCLEX exam is an example of a criterion-referenced exam.

Item creation. The researcher reviewed the literature from a variety of disciplines and compiled the list of best practices. When no new practices were identified, the researcher determined that saturation had occurred. The best practices compiled list is located in Appendix A. Beyond a description of pre-exam planning and blueprint development, the list includes general guidelines for item creation as well as specific rules that apply to the item stem and the distractors. (Coffman et al., 2010; Cox, 2019; D' La & Visbal-Dionaldo, 2017; Ibrahim, 2019; Khafagy et al., 2016; Killingsworth et al., 2015; Naeem et al., 2012; National League for

Nursing, 2012; Oermann & Gaberson, 2021; Ray et al., 2018; Rudolph et al., 2019; Schroeder, 2013; Tarrant & Ware, 2012).

Pre-test review. Once the items are chosen, best practices require a thorough proofread of the exam as a whole along with a peer review of items. Peer reviewers who have experience with creating and revising exam items should be asked to review the exam prior to administration. Peer review promotes accuracy. Proofreaders should ensure that answers or data in a stem do not provide clues to other items within the exam and read for understandability, screening for writing flaws. Directions for the exam should also be reviewed for clarity and conciseness (Cox, 2019; Killingsworth et al., 2015; Moore, 2020; Naeem et al., 2012; National League for Nursing, 2012; Oermann & Gaberson, 2021; Rudolph et al., 2019; Schroeder, 2013; Tarrant & Ware, 2012).

Post-test review. Exam results should be analyzed by a software system to identify the difficulty index, discrimination index, distractor frequency distribution, distraction discrimination, and exam reliability. Statistics should be used to ensure fairness of exams and adjustments made to students' scores accordingly (Cox, 2019; D' La & Visbal-Dionaldo, 2017; Killingsworth et al., 2015; Mager et al., 2017; Naeem et al., 2012; National League for Nursing, 2012; Obon & Rey, 2019; Oermann & Gaberson, 2021)

Test item revision. All test items need editing and refinement and faculty should use the results of the statistical analysis to revise items prior to the next administration of the exam. Items with difficulty and discrimination indexes which do not meet the standards set by the testing policy should be examined. Nursing educators can increase difficulty and discrimination indices by revising items to better evaluate higher-order thinking. Strengthening non-functioning distractors also can improve an item's difficulty and discrimination index (Khafagy et al., 2016;

Killingsworth et al., 2015; Naeem et al., 2012; Obon & Rey, 2019; Oermann & Gaberson, 2021; Rudolph et al., 2019; Tarrant & Ware, 2012).

Development of Clinical Judgment

Nurses use clinical judgment to provide safe, effective care for increasingly complex patients. Betts et al. (2019) defined clinical judgment:

> Nursing clinical judgment is the observed outcome of critical thinking and decision-making. It is an iterative process that uses nursing knowledge to observe and assess presenting situations, identify a prioritized client concern, and generate the best possible evidence-based solution in order to deliver safe client care. (p. 23)

Betts et al. (2019) discussed the importance of designing multiple-choice items that promote the development and assessment of clinical judgment. Using Benner's (1984) theory of Novice to Expert, Betts et al. (2019) recognized that nursing graduates should be advanced beginners, and therefore assessment of nursing students should be scaffolded throughout the curriculum to develop clinical judgment. As students progress through the curriculum, exam items should progressively test clinical judgment development to match the student level (Eweda et al., 2020). The item creation process described by Betts et al., (2019) emphasized the importance of best practices, peer review, posttest statistical analysis, and item revision.

Barriers to Exam Item Best Practice Implementation

Researchers identified several barriers to the use of exam item best practices such as a knowledge gap, lack of faculty development, time constraints, budget constraints, lack of policies to support best practices, and insufficient administration and administrative support (Birkhead et al., 2018; Halstead, 2013; Ibrahim, 2019; Lavin & Rosario-Sim, 2013; Moore, 2020; Oeon & Rey, 2019; Ray et al., 2018; Tarrant & Ware, 2012). The number and

combinations of barriers could prevent faculty from utilizing exam item best practices. As previously discussed, as content experts nursing faculty may lack the knowledge and training to create and analyze items (Ray et al., 2018). Quality item creation demands time and comprehensive item analysis requires software (Cox, 2019; Oermann & Gaberson, 2021; Rudolph et al., 2019; Schroeder, 2013). Faculty who work within departments without a testing policy, without exam software, or without support for faculty such as administrative personnel to help with the formatting and production of tests may experience further obstacles to implementation (Birkhead et al., 2018; Halstead, 2013; Ibrahim, 2019; Lavin & M Rosario-Sim aria, 2013; Moore, 2020; Obon & Rey, 2019; Ray et al., 2018; Tarrant & Ware, 2012).

Facilitators to Exam Item Best Practice Implementation

Potential facilitators for utilization of exam item best practices include faculty development, the use of exam blueprints, clear detailed policies, department item developer position to support faculty, peer feedback mechanisms, mentoring for new faculty, and administration and administrative support (Birkhead et al., 2018; Halstead, 2013; Ibrahim, 2019; Lavin & Rosario-Sim, 2013; Moore, 2020; Obon & Rey, 2019; Ray et al., 2018; Tarrant & Ware, 2012). Moore (2020) found that the more experienced faculty have been creating, analyzing and revising test items, the more proficient they become. Faculty need experienced peers and practice to improve their assessment abilities. Administration leaders can facilitate a culture and environment conducive to learning and promoting faculty development (Josiah Macy Jr. Foundation, 2018; Morrill, 2010; Witherspoon, 1997). With these elements in place, faculty will more easily adopt best practices of item creation and analysis.

Summary

Researchers and nursing educator experts have identified best practices to ensure exam construction. Faculty need instruction, tools, and support that help them utilize these practices. Cornerstones of best practice include the use of blueprints in exam creation with posttest item analysis. Faculty should revise items based on the posttest item analysis (Birkhead et al., 2018; Cox, 2019; D' La & Visbal-Dionaldo, 2017; Ibrahim, 2019; Libner & Kubala, 2017; Moore, 2020; Obon & Rey, 2019; Tarrant & Ware, 2012). In this study, faculty knowledge and use of exam item creation best practices were both examined along with faculty identified facilitators and barriers to using best practices in exam item and test construction.

Chapter 3: Methodology

Introduction

In order to gather data regarding nursing faculty's use of exam item best practices and potential facilitators and barriers to implementation, the researcher designed a Qualtrics survey that was disseminated via email. Obtaining the data from a cross-section of faculty from a variety of colleges and locations has the potential to gather a geographically diverse sample and improve the generalizability of the results (Orcher, 2014).

Conceptual Framework

A conceptual framework offers a foundation for the design of research, including the identification of key ideas, variables, and constructs (Roberts, 2010). The Assessment Design Decisions Framework (Bearman et al., 2014a) serves as the conceptual framework for this research. The framework provides educators with a systematic method of improving assessment without being prescriptive and allowing for differences between subject matter (Bearman et al., 2016). The framework consists of six categories arranged as a circle to eliminate the notion of supremacy and invite the educator to engage with the elements independently and within the whole. The six categories include purposes, context, learner outcomes, tasks, feedback processes, and interactions (Bearman et al., 2014b). See Figure 1 for a visual representation of the categories. Each category of the framework will be discussed in the following paragraphs.

Figure 1. Assessment Decision Making Framework (Bearman et al., 2014a)

    Educators use assessment activities for multiple purposes. Bearman et al. (2014b) identified three general functions of assessment: support learning, generate grades and impact future thinking. Educators create assessments to support student learning to ensure achievement of key learning outcomes, guide student engagement with content to prompt the learning desired and provide the students with feedback to strengthen the future application of learning (Bearman et al., 2014b; Bearman et al., 2016). Educators may label assessments to support learning as formative assessments. Educators also administer assessments to generate grades, also known as summative assessments. Grades provide a summary of student achievement of the learning outcomes for the course (Bearman et al., 2014b). The third purpose of assessment involves the development of student thinking. A well-crafted assessment not only supports learning and

generates grades, but also builds the students' ability to make judgments that apply to the task at hand and promotes judgment development (Bearman et al., 2014b). This type of assessment has been labeled sustainable assessment. Boud and Soler (2015) defined sustainable assessment as an activity that extends beyond the immediate task and develops student judgment. The ability to examine their own thinking processes allows the learner to self-evaluate and self-regulate, which assists the learner in identifying gaps, seeking assistance, and applying feedback (Bearman et al., 2014b). Therefore, the purpose of assessment encompasses the promotion of student learning, the generation of a grade, and the development of learner judgment.

Assessments occur within various contexts. The context of assessment contains various environmental and personal variables that influence assessment construct (Bearman et al., 2016). These variables include the characteristics of the learners, the institution's assessment policies, requirements related to professional and vocational accreditation, departmental expectations, the placement of the module within a course and where the course falls within the curriculum, and the learning environment (Bearman et al., 2014a). When designing assessments, educators must consider who the students are, where they are in the course of study, the type of learning environments, as well as the influence of institutional, departmental, and professional policies and expectations.

Educators align assessments with desired learner outcomes. Not only do assessments evaluate if students achieved the learner outcomes for the current unit, module, or course; but also aim to meet the overall program outcomes and professional requirements, plus develop general learner intellectual development (Bearman et al., 2014b). Hence, educators must consider a myriad of outcomes when creating assessments.

Assessments consist of various types of tasks. The educator chooses a task to match the type of assessment. Examples of formative assessment tasks include worksheets, sets of problems, case studies, and quizzes. Tasks that could correlate with summative assessment are multiple-choice tests, projects, and papers. Optimally designed assessment tasks also develop the learner's future judgment skills and remain sustainable from the perspective of the educator's workload (Bearman et al., 2014b). As educators design assessment tasks, they analyze how the tasks fit holistically into the learner outcomes.

Learners require feedback when they prepare for assessment and after a completed assessment in order to correct misconceptions and encourage the further application of learning. In order for feedback to be the most effective, students should receive multiple opportunities throughout a course to complete assessments, receive feedback, and demonstrate an understanding of feedback (Bearman et al., 2014b). Feedback need not be given only by the educator but may involve peers, tutors, or resources with online learning systems.

Interactions form a variable within the Assessment Design Decisions Framework. One aspect of interaction involves the communication process between the educator and the student, as well as between the educator and department colleagues and/or other stakeholders. This communication promotes engagement in and collaboration on assessments. Interactions also include how the assessment connects with the other teaching and learning elements of the course; how this piece fits into the whole (Bearman et al., 2014b).

The elements of the Assessment Design Decision Framework guide the research for the use of best practices for exam item creation. The variables within the framework provide the foundational rationale for why nursing faculty need to implement best practices to improve assessment. By keeping the purposes, context, learner outcomes, tasks, feedback processes, and

interactions in mind, faculty can ensure that assessment participates in reaching the ultimate end goal of nursing education and graduates individuals who practice clinical judgment in order to best care for patients and ensure patient safety (Dickison et al., 2019; Oermann & Gaberson, 2021).

Research Design

The research was completed by using a quantitative cross-sectional design with data collection via a Qualtrics survey. The Qualtrics survey was sent to nursing faculty across the United States. Once collected, data analysis was completed using SPSS to run a variety of data analyses to classify variables and identify variable associations.

Research Questions

RQ1: What relationship, if any, exists between faculty's demographic factors (i.e., age, gender) and their score on the exam item best practices list?

H_o1: There is no relationship between faculty's demographic factors and their utilization of exam item best practices score.

RQ2: What relationship, if any, exists between faculty's individual factors (i.e., level of education, length of time since highest degree completion, years of teaching experience, frequency of writing exam items, certification as Certified Nurse Educators (CNE)) and their score on the exam item best practices list?

H_o2: There is no relationship between faculty's individual factors and their utilization of exam item best practices score.

RQ3: What relationship, if any, exists between faculty's environmental factors (i.e., type of institution at which the participant works, department has a testing policy, administrative

support available, administration support of exam creation best practices, and availability of funds for faculty development) and their score on the exam item best practices list?

H_o3: There is no relationship between faculty's environmental factors and their utilization of exam item best practices score.

RQ4: What relationship, if any, exists between faculty's social factors (i.e., formal and/or informal mentor) and their score on the exam item best practices list?

H_o4: There is no relationship between faculty's social factors and their utilization of exam item best practices score.

RQ5: What is the frequency and variation of ratings of the faculty-identified barriers and facilitators of exam item best practice implementation?

RQ6: What associations exist between faculty-identified barriers and facilitators of exam item best practice implementation and faculty's demographic, individual, environmental, and social factors?

Sample

Nursing faculty currently working at schools of nursing that offer entry-level nursing programs comprised the proposed sample for this study. The American Association of Colleges of Nursing maintains a directory of 878 nursing programs across the United States. The directory contains links to the individual programs' websites. The number of member schools was narrowed to 744 by choosing programs accredited by the Commission on Collegiate Nursing Education (CCNE) (American Association of Colleges of Nursing, 2021a). CCNE accreditation is given to baccalaureate and graduate nursing programs (American Association of Colleges of Nursing, 2021b). Some nursing departments also may house two- or three-year associate's programs. Four faculty who teach in an associate's degree program were reached and

participated in the survey. Email addresses of 510 nursing deans, chairs, or directors were collected by exploring the 744 identified links of the programs identified as having an entry-level registered nurse program. The emails sent to the leadership faculty requested the leaders to distribute the survey to their faculty.

Colleagues within the researcher's department offered to share contact information for nursing faculty with whom they have personal connections to hopefully increase participant response. These colleagues also were invited to complete the survey ($N = 43$). With a wide variety of access points to different nursing departments, the researcher hoped to reach faculty who work in entry-level nursing programs across the United States. The potential participants were drawn from a wide range of geographical locations and a variety of institutional types. The pool qualifies as a convenience sample because the researcher sent the survey to faculty whose email addresses were obtained; not all the possible nursing faculty in the United States (Creswell, 2014). As an incentive to complete the survey, participants could choose to enter a drawing for one of ten $75 Amazon.com gift cards and/or receive an executive summary of the research.

Instrumentation

The Qualtrics survey design was based upon the review of the literature. The survey contained three sections and was emailed to nursing department leadership who forwarded it to the nursing faculty. The first part of the survey focused on the knowledge and utilization frequency of exam item best practices. A list of best item practices identified in the literature was presented to participants in a grid format. Participants indicated via a Likert scale if they are: 0 = not familiar with the practice, 1 = never use the practice, 2 = sometimes use the practice, 3 = frequently use the practice, or 4 = always use the practice. A "score" for best practice use

was calculated upon the completion of this section. A higher score indicated a consistent and broad use of best practices. The participant's score constituted the dependent variable.

The second section of the survey presented participants with the barriers to implementation of exam item best practices and the facilitators of implementation identified in the literature. The barriers were presented in a grid with answer options: Yes = agree this is a barrier to implementing exam item best practice, No = this is not a barrier to implementing exam item best practice. The facilitators found in the literature were individually listed and participants were asked to choose what they believed were the top three facilitators of exam item best practice implementation. After completing the questions regarding barriers and facilitators, the participant was given an opportunity to enter an "other" answer if they identified a barrier or facilitator to the implementation of exam item best practices that were not included in the survey. The answers given in the "other" sections provided insight regarding additional implementation barriers or facilitator participants identified.

The third and final section collected faculty variables including demographic, individual, environmental, and social variables. Each of the faculty variables was chosen based on previous research reported in the literature (Birkhead et al., 2018; Cox, 2019; Ibrahim, 2019; Killingsworth et al., 2015; Lavin & Rosario-Sim. 2013; Libner & Kubala, 2017; Naeem et al., 2012; O'Rae et al., 2019; Ray et al., 2018; Tarrant & Ware, 2012). The faculty factors were the independent variables.

Demographic variables included the age and gender of the participants. For the individual variables, participants were asked to answer questions regarding the highest level of education achieved, the number of years since the highest degree was completed, the number of years of experience they have accumulated, how frequently they write exam items, and whether

they have earned the certification of Certified Nurse Educator (CNE). The participant indicated their highest level of education achieved, choosing from the various possibilities. The questions regarding their number of years of experience, the frequency of writing exam items, and number of years of experience had a numerical range from which the participant chose. The CNE question was in yes/no format.

Questions that correlate with environmental variables inquired about the type of institution at which the participant works, in which type of program the participant teaches, whether or not their department has a testing policy, if the participants have any administrative support available to help with the clerical tasks of exam creation (e.g., uploading items into the learning management system), if the administration strongly supports exam creation best practices, if the institution at which they work provides funds for faculty development, and if their department utilizes a software program for exam item statistics. The participant chose from a list of the institution and program types. The other environmental variable questions were in yes/no format.

The last question in the faculty variable section inquired if participants have a formal or informal mentor within their department. The participants selected from no mentor, formal mentor, or informal mentor who assists you with test writing. The participant could choose both an informal and formal mentor.

Participants needed approximately 15 minutes to complete the survey. The survey contained three sections of questions, many of which were housed with a grid to be answered via a Likert scale or yes/no format. Some demographic questions were included at the end of the survey.

Pilot test.

Several retired nursing faculty with past experience in exam item creation and in research agreed to take the survey as a pilot test once IRB approval was obtained. This provided the researcher with the opportunity to receive feedback regarding the survey and establish an estimated time to complete the survey. Feedback from the participants contributed to the rewording of several faculty variables for the sake of clarity.

Data Collection

Surveys were sent to nursing faculty after email addresses were obtained. The email invited participants to participate in an online survey regarding best practices in exam item creation and revision. If the participants chose to participate, an embedded link in the email brought them to the Qualtrics survey. The first page of the survey contained the informed consent. The informed consent included sections addressing: the study's topic and purpose, the participants, voluntary participation, confidentiality, participant risks, benefits to participants, future use of data, the researcher's contact information, and the institutional review board's approval of the project. Once the participant electronically signaled their agreement, the survey progressed to the sections described above.

The survey was sent to heads of nursing departments of potential participants via email. The email asked that the department leader forward the email to nursing faculty within their department. A link in the email brought the participants to the Qualtrics survey. The informed consent included the study's topic and purpose, a description of those invited to participate, a statement regarding voluntary participation, survey, protocol, how confidentiality will be maintained, and a description of participant risks and potential benefits. The informed consent also described any future use of data and how to contact the researcher. Finally, the IRB

approval of the study was described. The participant was asked to verify that they read the informed consent page and agreed to participate in the study. A statement regarding their ability to skip questions or not complete the survey was also included. A follow-up email to participants was sent two weeks and four weeks after the survey encouraging faculty to complete the questionnaire.

Data Analysis

Before the data were collected, a multiple regression was the proposed analysis for research hypotheses #1, #2, and #3. These research questions examined the statistical relationships between a participant's score on the exam creation best practices list and demographic, individual, and environmental faculty factors. Question #1 addressed faculty demographical data regarding age and gender. Question #2 explored individual faculty data regarding the highest level of education, the number of years since highest degree completion, years of teaching experience, frequency of writing exam items, and if the participant has certification as a Certified Nurse Educator (CNE). Question #3 requested participants' information regarding environmental factors, such as the type of institution in which they work, if their department has a testing policy, if there is administrative support available, if their administration supports exam creation best practices, and if the participants receive funds for faculty development. These factors do not exist in isolation and multiple linear regression provided information about the strength of the relationships among the faculty variables (independent variables) and the score on the exam creation best practices list (dependent variable) (Muijs, 2011).

The proposed analysis for the participant data gathered in research hypothesis #4 was a one-way ANOVA. The one-way ANOVA allowed comparison between a group and the score

46

on the exam item creation best practices list (Muijs, 2011). Research question #4 groups participants by the social factor of mentorship that assists with exam item creation, evaluation, and revision. Participants fell into three groups: no mentor, informal mentor, formal mentor.

Research questions #5 gathered data regarding what faculty identified as barriers to and facilitators of the use of exam item creation best practice (#5). These data were analyzed using descriptive statistics to summarize faculty responses and identify potential patterns (Muijs, 2011).

A series of Chi-square tests were proposed to analyze the hypotheses related to research question #6. The data from research question #6 identified associations between what faculty classified as barriers to and facilitators of exam item best practices and the discrete variables identified as faculty factors (e.g., demographic, individual, environmental, social factors).

Table 1

Research Questions and Data Analyses Methods

Research questions	Proposed Data analysis
RQ1: What relationship, if any, exists between faculty's demographic factors (i.e., age, gender) and their score on the exam item best practices list? RQ2: What relationship, if any, exists between faculty's individual factors (i.e., level of education, length of time since highest degree completion, years of teaching experience, frequency of writing exam items, certification as Certified Nurse Educators (CNE)) and their score on the exam item best practices list? RQ3: What relationship, if any, exists between faculty's environmental factors (i.e., type of institution at which the participant works, department has a testing policy, administrative support available, administration support of exam creation best practices, and availability of funds for faculty development) and their score on the exam item best practices list?	Multiple linear regression
RQ4: What relationship, if any, exists between faculty's social factors (i.e., formal and/or informal mentor) and their score on the exam item best practices list?	One-way ANOVA
RQ5: What is the frequency and variation of ratings of the faculty-identified barriers and facilitators of exam item best practice implementation?	Descriptive statistics
RQ6: What associations exist between faculty-identified barriers and facilitators of exam item best practice implementation and faculty's demographic, individual, environmental, and social factors?	Chi-square test

Limitations/ Delimitations

Surveys rarely achieve a high participation rate (Orcher, 2014). One inherent limitation with research via surveys is that respondents may be persons who are vested in the survey topic (Orcher, 2014). Nursing faculty who hold a greater interest or investment in exam item creation may be more inclined to answer the survey. Faculty who feel ill-equipped in the area of exam item creation or do not regularly write exam items in their practice may choose not to participate in the survey. Either scenario will influence the reliability of the data gathered. Studies show that lottery incentives increase the likelihood of participation (Laguilles, Schweiger, & Sapienza, 2011). Offering participation in a drawing for an Amazon.com gift card may incentivize participants to complete the survey. A follow-up email to participants was sent twice during the four-week period that the survey was open encouraging faculty to complete the questionnaire.

As mentioned previously, although several researchers and assessment experts from higher education have suggested activities that best guide the creation, analysis, and revision of exam items (Ibrahim, 2019; Killingsworth et al., 2015; Naeem et al., 2012; National League for Nursing, 2012; Oermann & Gaberson, 2021; O'Rae et al., 2019; Rudolph et al., 2019), one defined list of best practices have not been adopted by nursing education organizations. This lack of national nursing education support could act as a deterrent to implementation for some participants who receive the survey.

The choice to email the survey to the nursing faculty for whom the researcher could obtain addresses may be seen as a delimitation. Exclusion of nursing faculty may occur secondary to the availability of email addresses, which limited the pool of potential respondents and may impact the validity of data secondary to which faculty receive the survey. The nursing

leaders acted as a gatekeeper and, if they lacked the time or resources to forward the email to faculty, the number of potential participants was decreased.

Ethical Issues

This study held minimal risk for the participants. Participants' identity was not linked to their responses on the survey. The survey did not collect the participant's name, email address, or exact locations of participants. In order to sign up for the drawing for the Amazon.com gift card and/or an executive summary of the research, the participant followed a link to a google form, outside of the survey, to provide their name and email address. The information on the google form was not linked to the responses within the survey. Participants' identities remained anonymous within the survey. The names and email addresses of the participants who requested to participate in the drawing and/or the executive summary were deleted once the study was complete. One possible risk involved eliciting negative participant feelings when taking the survey (Merriam & Tisdel, 2016; Orcher, 2014; Patten, 2014). An example of negative feelings includes guilt for not implementing best exam practices. Participants were instructed that they could skip questions or exit the survey at any time.

A further ethical consideration focuses on the benefit of the study beyond the researcher. A subtle way of exploiting participants includes not providing some type of reciprocity. Although researchers may be personally interested in a specific topic, a study should benefit participants and/or others, not only the researcher (Creswell, 2014). Reciprocity need not be a grand gesture but may be minor, such as the dispersion of results to the participants (Bogdan & Biklen, 2007; Creswell, 2014; Orcher, 2014; Patten, 2014). The researcher offered participants the opportunity to receive a report of the study findings to provide a level of reciprocity.

The study aligned with the statutes established in the Belmont Report (Department of Health, Education, and Welfare, 1979). Through the use of informed consent, steps were taken to ensure participants received information regarding possible risks and emphasized the voluntary participation and option to withdraw from the survey at any point during survey completion.

Chapter 4: Results

Introduction

The implementation of exam item best practices increases the quality of exam items used in nursing education and better prepares students for taking the board exam and for entry-level practice. The first purpose of this study was to examine relationships between nursing faculty's demographic, individual, environmental, and social factors and their knowledge and use of exam item best practices. The second purpose was to examine what nursing faculty identified as potential barriers and facilitators of exam item best practices implementation.

Variables

The dependent variable for research questions one through four was the participant's score on the exam item best practice list. Participants were asked to rate how frequently they use the exam item best practices in pre-exam planning, general item creation, creation of item stems, creation of item distractors, pre-exam review, post-exam analysis, and item revision. Response choices included: not familiar with the practice, do not use, sometimes, half of the time, most of the time. If participants were unfamiliar with the practice, they were asked to choose "not familiar with the practice" rather than "do not use" For scoring on each practice, if the participant selected "not familiar with the practice" or "do not use" they received "0", "sometimes" received "1", "half of the time" received "2 ", and "most of the time" received "3". The participant's score was calculated by the average score of all best practices used. A perfect score would equate 3.0. Participants who did not complete at least 70% of the best practices did not receive an exam best practice score. Their exam best practice score (EBP score) was the dependent variable for research questions one, two, three, and four.

The independent variables for research questions one through four were the faculty demographic, individual, environmental, and social variables. Please see Table 3 for the variables within each category.

Research question five looked at the frequency and variation of ratings of the faculty-identified barriers and facilitators of exam item best practice implementation. The list of potential barriers and facilitators identified in the literature and included in the survey can be found in Table 15 (*Ranks and Frequencies for Barriers to Best Practices*) and Table 17 (*Ranks and Frequencies for Facilitators of Best Practices*).

Research question six examined the associations that exist between faculty-identified barriers and facilitators of exam item best practice implementation and faculty's demographic, individual, environmental, and social factors.

Sample

The potential number of participants was not known. The survey was sent to 510 heads of nursing departments of potential participants via email. The email asked that the department leader forward the email to nursing faculty within their department. A link in the email brought the participants to the Qualtrics survey. The number of potential participants was unknown because the number of faculty within each department varied and nursing leaders could choose to send the email to all faculty or to those who may be interested in the survey. Two follow-up emails were sent to the nurse leaders two and four weeks after the initial email. The invite to participate in the survey was also posted on the American Association of Colleges of Nursing's community member board twice. The researcher completed three institutions' IRB applications and shared her institutional IRB approval upon request.

The sample size varied for each question. Participants were allowed to skip a question if desired, therefore incomplete surveys were included in the data. Please see Tables 2 and 3 for the number of participants who answered the questions in the various sections of the survey.

Table 2

Completion Rate for Survey Sections

Section 1: Exam item best practices		Number of participants
	Pre-exam planning	307
	General exam creation	301
	Creation of item stem	296
	Creation of item distractors	292
	Pre-exam review	287
	Post-exam review	286
	Item revision	286
Section 2: Barriers and Facilitators		
	Barriers	284
	Facilitators	284
Section 3: Faculty demographics and variables		
	Demographic	285
	Individual	285
	Environmental	123-285 (range)
	Social	133

Table 3

Faculty Variables: Categories, Questions, and Potential Responses

Faculty variables category	Questions:	Potential response:	Total number of respondents per question
Demographic	What is your age?	24 or younger 25-34 35-44 45-54 Older than 55	285
	What is your gender?	Male Female	285
Individual	What is the highest level of education you have obtained?	Doctorate Master's Bachelor's Associate's	285
	How many years since you completed your highest degree?	0-2 years 3-5 years 6-10 years Greater than 10 years	285
	How many years of experience do you have teaching nursing?	0-2 years 3-5 years 6-10 years Greater than 10 years	285
	How frequently do you write exam items?	Weekly Monthly One to three times per semester Once to twice per year Never	285
	Are you a certified nurse educator?	Yes No	285
Environmental	At what type of institution do you work?	Private college or university Public university	285

		For-profit institution	
	Does your nursing program have a testing policy that includes the best practices for exam creation?	Yes No	124
	Do you receive administrative support for the creation of exams or tacking of exam results?	Yes No	123
	Does your nursing department use software to track exam statistics?	Yes No	124
	Does your institution provide funds for faculty development?	Yes No	124
	Does your department utilize exam blueprints?	Yes No	124
	Does your administration support faculty in the use of exam item best practices?	Yes No	124
Social	Do you have a nursing faculty mentor who helps you with exam item development?	Yes – formal mentor Yes – informal mentor No	133

Descriptive Statistics for Individual Exam Best Practices Items

Before addressing the six main research questions, means and standard deviations for the individual items from Exam Best Practices (EBP) scale were created (see Table 4). The items with the highest mean were: *Use appropriate vocabulary (avoid colloquialisms or slang terms);* *Write exam items free of grammatical and structural errors, good technical quality;* and *Protect the integrity of the exam.* The three items with the lowest means were; *Consider three-option items instead of 4-5 option items if an additional distractor is not plausible*; *Develop a blueprint as defined below; if your blueprint does not contain all the elements, please select not familiar with practice or do not use;* and *Specify the desired difficulty and discrimination level of the items.*

Table 4

Descriptive Statistics: Faculty Use of Exam Best Practices

Category: Exam Best Practice	N	Mean	Std. Deviation
General for Item Creation: Use appropriate vocabulary (avoid colloquialisms or slang terms)	286	2.95	.298
General for Item Creation: Write exam items free of grammatical and structural errors, good technical quality	286	2.93	.354
Post-Exam Review: Protect the integrity of the exam	284	2.91	.410
Pre-Exam Review Ensure instructions are concise, clear and not open to further explanation	285	2.88	.459
Item Distractor Ensure that all options are grammatically consistent with the stem	285	2.87	.478
Pre-Exam Review Proofread exam for understandability and conflicts between questions	285	2.86	.500
Item Distractor Verify that all options are mutually exclusive (e.g. number ranges do not overlap)	281	2.84	.527
General for Item Creation Ensure wording and sentence structure is succinct	286	2.84	.438
General for Item Creation Abstain from stereotyping race, gender, or other factors/Screen for offensive content or scenarios	286	2.82	.594
Pre-Exam Review Proofread to ensure that answers/clues to one item are not provided elsewhere within the exam	285	2.76	.603
General for Item Creation Write questions to test higher cognitive thinking	285	2.75	.530
Item Stem	284	2.74	.571

Write items that have only one correct answer (except in cases where 'select all' is specified)			
Item Distractor	285	2.72	.729
Distribute correct answers randomly and evenly			
Item Stem	286	2.67	.688
Explicitly state the information you are seeking; the question clearly defines the problem			
Pre-Exam Review	284	2.67	.773
Screen for writing flaws			
General for Item Creation	286	2.66	.837
Avoid hinged questions – questions that rely on answer from previous question			
Item Distractor	284	2.66	.656
Ensure that all options equal in length and amount of detail			
Post-Exam Review	285	2.66	.848
Utilize psychometric analysis to assure that the test is valid and internally consistent			
General for Item Creation	285	2.62	.714
Support questions, answers, and rationales with research and/or other reputable evidence			
Item Stem	285	2.60	.704
Avoid providing cues to the correct answer within the stem			
Post-Exam Review	283	2.59	.924
Track item discrimination index			
Item Distractor	285	2.59	.842
Avoid the use of "all of the above" or "none of the above"			
Item Stem	286	2.55	.688
Avoid extraneous material not needed to answer the question			
Item Distractor	284	2.55	.684
Evaluate that all options are plausible; options should be homogenous without obvious outliers in content			
Pre-Exam Planning	286	2.55	.892

Specify the number of items on the exam and the proportion of items per content area/objectives			
Item Distractor	285	2.54	.743
Avoid the use of vague terms (e.g. frequently, often, occasionally) in all the options			
General for Item Creation	285	2.54	.709
Avoid the use of absolute terms (e.g. always, never, all, only)			
Item Revision	285	2.53	.870
Use statistical data to revise item stems prior to next administration of exam			
General for Item Creation	285	2.52	.890
Avoid testing on material deemed trivial			
Item Stem	286	2.51	.776
Ensure that the stem can stand alone, is clear and complete (e.g. May be answered by competent student without provided choices)			
Pre-Exam Planning	285	2.51	.944
Plan for a sufficient number of items (50-60 items to achieve high level of reliability)			
Post-Exam Review	285	2.46	1.019
Track difficulty index			
Item Stem	286	2.38	.987
Write stem as a question or partial statement			
Item Revision	285	2.34	1.017
Use statistical data to revise item distractors prior to next administration of exam			
Item Stem	286	2.32	.883
Avoid the use of negative words in the stem (e.g. except, not incorrect)			
Pre-Exam Planning	286	2.27	1.112
Define exam purpose, congruent with module/course objectives			
Item Distractor	284	2.23	.837
Resist presenting false information via distractor items – use all correct answers, forcing students to pick the "best" answer			
Item Distractor	284	2.18	1.202

Arrange options in a logical order (e.g. numerical, alphabetical)			
Pre-Exam Planning	286	2.17	1.091
Specify the format of exam items, including number of alternative items			
Post-Exam Review	284	2.17	1.224
Track exam reliability (e.g. KR20)			
General for Item Creation	285	2.10	.968
Avoid use of items from commercial item banks			
General for Item Creation	286	2.07	1.013
Create a pool of high-quality test items (faculty generated)			
Post-Exam Review	284	2.02	1.234
Track distractor efficacy			
General for Item Creation	286	1.97	1.001
Use clinical vignettes/scenario based			
Pre-Exam Review	285	1.84	1.161
Corroborate with peer to complete review of exam items			
Item Stem	286	1.78	1.277
Emphasize key words (e.g. best, priority, first) in italics, underlined, or bolded			
Pre-Exam Planning	286	1.74	1.204
Specify the desired difficulty and discrimination level of the items			
Pre-Exam Planning	281	1.46	1.333
Develop a blueprint.			
Item Distractor	285	1.01	1.197
Consider three option items instead of 4-5 option items if an additional distractor is not plausible.			
Valid N (listwise)	262		

Research Question One

The focus of research question one was the relationship between the nurse faculty member's demographic factors (gender and age) and exam best practice (EBP) scores. The original independent variables, gender and age of the nurse educator, were both problematic. Gender cannot be used in the analyses because there were too few male nurses (N = 9). The age variable needed to be recoded because of the fewer number of younger nurses. Table 5 has the original frequency distribution. For the recoded variable, younger than 24, 25-34 years old, and 35-44 years old categories were combined into one category labeled "44 years old and younger."

Table 5

Original Frequency Distribution for Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	< 25	0	0.0	0.0	0.0
	25-34	21	7.3	7.4	7.4
	35-44	47	16.3	16.5	23.9
	45-54	76	26.4	26.7	50.5
	Older than 55	141	49.0	49.5	100.0
	Total	285	99.0	100.0	
Missing	System	3	1.0		
Total		288	100.0		

A one-way ANOVA was used to analyze EBP Score by the age of the nurse educator. First, Levene's statistic was used to test the homogeneity of the variances assumption. The assumption was not violated ($p = .129$). The overall one-way ANOVA revealed a significant difference between the means, $F(2, 282) = 6.12$, $p = .003$. Post hoc tests revealed that younger nurse educators (44 years old and younger) had significantly lower EBP scores ($M = 2.36$) compared to older nurse educators (45-54 years old, $M = 2.53$; 55 and older, $M = 2.54$).

Table 6

Descriptive Statistics for EBP Score by Age

	N	Mean	Std. Deviation	Std. Error
44 and younger	68	2.3645	.44390	.05383
45-54	76	2.5268	.29372	.03369
55 and older	141	2.5368	.32121	.02705
Total	285	2.4930	.35404	.02097

Table 7

One-way ANOVA for EBP Score by Age

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.481	2	.740	6.119	.003
Within Groups	34.117	282	.121		
Total	35.598	284			

Table 8

Post Hoc Tests for EBP Score by Age

(I) Age	(J) Age	Mean Diff. (I-J)	Std. Error	Sig.
44 and younger	45-54	-.16230*	.05806	.006
	55 and older	-.17233*	.05135	.001
45-54	44 and younger	.16230*	.05806	.006
	55 and older	-.01003	.04950	.840
55 and older	44 and younger	.17233*	.05135	.001
	45-54	.01003	.04950	.840

*. The mean difference is significant at the 0.05 level.

Research Question Two

       Research question two focuses on the relationship between the faculty member's individual factors and EBP score. As in research question one, some of the original independent variables needed to be recoded. Level of education (Degree) was recoded to a dummy variable with 0 = Master's and 1 = Doctorate. As can be seen in Table 9, the "How frequently do you write exams?" variable needed to have the three "Never" categories combined together. It was recoded as follows: 1 = Never, 2 = Once to twice per year, 3 = One to three times per semester, 4 = Monthly, 5 = weekly. The item "Are you a Certified Nurse Educator (CNE)?" was dummy coded with 0 = No and 1 = Yes. The years of teaching variable was recoded with "less than 10 years" = 0 and "10 or more years" = 1. The years since completed highest degree variable was kept the same.

Table 9

Frequency Distribution for the item "How frequently do you write exam items?"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Weekly	44	15.3	15.4	15.4
	Monthly	109	37.8	38.2	53.7
	One to three times per semester	93	32.3	32.6	86.3
	Once to twice per year	32	11.1	11.2	97.5
	Never - standardized curriculum or program does not use tests	2	.7	.7	98.2
	Never - use commercial text bank questions without revision	3	1.0	1.1	99.3
	Never - department has exam item writer	2	.7	.7	100.0
	Total	285	99.0	100.0	
Missing	System	3	1.0		
Total		288	100.0		

A multiple linear regression was used to see if any of the independent variables (listed above) were significant predictors for the EBP Score. Three of the variables were significant predictors of EBP Scores. First, the more years of experience teaching, the higher the EBP Score, $\beta = .246$, $p < .001$. Next, the more frequently the nurse educator wrote exams, the higher their EBP Score, $\beta = .246$, $p < .001$. Finally, nurses who were CNE's had significantly higher EBP Scores, $\beta = .143$, $p = .014$. See Table 10 for more details on predictor variables.

The overall model was statistically significant, $F(5,279) = 11.41$, $p < .001$, $R^2 = .170$. That means that the predictor variables can account for 17% of the variance in EBP Scores.

Table 10

Betas and Significance Values for Individual Factors (Predictors) of EBP Scores

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.062	.095		21.691	.000
	Degree	-.077	.045	-.105	-1.732	.084
	Yrs Teaching	.177	.049	.246	3.591	.000
	Frequency of Exam Writing	.089	.020	.242	4.388	.000
	CNE	.111	.045	.143	2.466	.014
	Years since you completed highest degree	.011	.021	.032	.506	.614

a. Dependent Variable: EBP Score

Table 11

Model Summary for Multiple Regression of Individual Factors and EBP *Scores*

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.412[a]	.170	.155	.32547

a. Predictors: (Constant), How many years since you completed your highest degree?, CNE, Frequency of Exams, Degree, Years Teaching

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.043	5	1.209	11.410	.000[b]
	Residual	29.555	279	.106		
	Total	35.598	284			

a. Dependent Variable: BestPracticesScore
b. Predictors: (Constant), How many years since you completed your highest degree?, CNE, FrequencyofExams, Degree, YrsTeaching

Research Question Three

The third research question focused on the relationship between the faculty's environmental factors (i.e., type of institution at which the participant works, the department has a testing policy, administrative support available, administration support of exam creation best practices, and availability of funds for faculty development) and their score on the exam item best practices list. All independent (or predictor) variables were dummy coded with 0 = no and 1 = yes, except for the institution type variable which was coded as 0 = Private college or university and 1 = Public university. There were two few participants from "For Profit" universities to be included in the analyses (N = 4).

A multiple regression analysis was run with the independent variables (listed above) and the EBP Score as the dependent variable. None of the predictors were significantly related to the

EBP Score (see Tables 12 and 13 for detailed results). The overall model did not quite reach statistical significance, $F(7,110) = 1.975$, $p = .065$, $R^2 = .112$.

Table 12

Betas and Significance Results for Environmental Factors and EBP *Scores*

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.256	.107		21.001	.000
	TestingPolicy	.133	.085	.154	1.565	.121
	Blueprint	-.170	.093	-.196	-1.832	.070
	Software	.046	.077	.056	.599	.550
	AdminSupport	.182	.109	.169	1.674	.097
	LeaderSupport	-.029	.075	-.037	-.386	.700
	FacDevFunds	.164	.090	.170	1.825	.071
	TypeofInstitution	.123	.072	.158	1.720	.088

a. Dependent Variable: BestPracticesScore

Table 13

Model Summary for Multiple Regression of Environmental Factors and EBP *Scores*

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.334[a]	.112	.055	.37932

a. Predictors: (Constant), TypeofInstitution, LeaderSupport, Software, FacDevFunds, AdminSupport, TestingPolicy, Blueprint

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.989	7	.284	1.975	.065[b]
	Residual	15.827	110	.144		
	Total	17.816	117			

a. Dependent Variable: BestPracticesScore
b. Predictors: (Constant), TypeofInstitution, LeaderSupport, Software, FacDevFunds, AdminSupport, TestingPolicy, Blueprint

Research Question Four

The fourth research question focused on the relationship between the faculty's social factors (formal mentor, informal mentor, and no mentor). A one-way ANOVA was conducted to determine if there was a relationship between the faculty social variable and EBP score. There were no significant differences between those with formal mentors, informal mentors, and no mentors on EBP, $F(2, 121) = 1.293$, $p = 0.278$. The means were essentially the same (formal mentor: m = 2.5, SD = 0.4; informal mentor: m = 2.6, SD = 0.3; no mentor: 2.4, SD = 0.4).

Table 14

Mentor Group Comparisons on EBP

Mentor	N	Mean	SD	95% Lower Bound	95% Upper Bound	F	p-value
Formal Mentor	15	2.5	0.4	2.3	2.7	1.293	0.278
Informal Mentor	40	2.6	0.3	2.5	2.6		
No Mentor	69	2.4	0.4	2.5	2.5		
Total	124	2.5	0.4	2.6	2.6		

Research Question Five

The fifth research question looked at the frequency and variation of ratings of the faculty-identified barriers and facilitators of exam item best practice implementation.

Barriers to Best Practices Participants were asked to respond 'yes' or 'no' to barriers identified in the literature. As seen in Table 15, faculty identified the biggest barrier to the use of best practices as time constraints, which was selected by 84.6% of the faculty. The second biggest barrier was lack of faculty development regarding item revision (69.4%), and the third biggest barrier was lack of educational preparation (63.7%). After the top three barriers, nurses selected lack of administrative support (61.9%), lack of faculty development regarding item creation (61.1%), lack of faculty development funds (59.1%), lack of faculty development regarding post-test review (56.0%), lack of testing policy (52.0%), lack of leader/administration support (50.7%), lack of peer support (42.3%), and lack of exam software (39.0%).

Table 15

Ranks and Frequencies for Barriers to Best Practices

RANK	Barrier	Yes	N	Percentage
1	Time constraints	237	280	84.6
2	Lack of faculty development re: item revision	179	284	69.4
3	Lack of educational preparation	179	281	63.7
4	Lack of administration support	174	281	61.9
5	Lack of faculty development re: item creation	173	283	61.1
6	Lack of faculty development funds	166	281	59.1
7	Lack of faculty development re: post-test review	168	282	56
8	Lack of testing policy	146	281	52
9	Lack of leader (administration support)	143	282	50.7
10	Lack of peer support	119	281	42.3
11	Lack of exam software	110	282	39

Participants were given an opportunity to type in a barrier if they felt something was not included in the survey question. Twenty participants entered additional barriers to best practices. Although lack of educational preparation and faculty development for item creation, posttest analysis, and item revision were included as barriers in the survey, 35% of the participants' comments contributed to the theme of lack of educational preparation and faculty development. The second theme of lack of peer support also emerged in 30% of the comments. And, a learning management system (LMS) that lacked the capability to support alternative testing items and gather desired statistics constructed the third theme with 10% of the responses. See Table 16.

Table 16

Additional Barriers to Best Practices

Rank	Percentage	Theme	Examples of comments
1	35%	Lack of educational preparation and/or faculty development	The inconsistent pedagogical preparation of academic nurse educators Absolutely no training when I started Requires ongoing faculty development
2	30%	Lack of peer/colleague support	Peer resistance to reviewing/critiquing exams Lack of collaboration among faculty to identify cognitive levels/elevation of testing skills aligned with complexity of courses Lack of faculty interest in exam item best practice
3	10%	LMS barriers	Lack of LMS support of alternative testing modalities LMS currently in use provides statistics that are irrelevant and difficult to understand

Facilitators of Implementation of Best Practices. Faculty were asked to identify what they perceived as the top three facilitators of best practice use from a list of facilitators found in current research. As seen in Table 17, faculty identified the top facilitator as faculty development regarding exam item best practices with 59.5%. The second highest-rated facilitator was exam item mentoring for new faculty (46.7%), and the third top facilitator was faculty within the department with expertise in item creation, exam evaluation, and item revision (37.7%). After the top three facilitators, faculty selected testing policy which includes exam creation best practices (32.5%), administration support for use of exam item best practices (26.3%), administrative support for exam creation and revision (24.9%), department item developer position to support faculty (23.5%), available faculty development funds (19.7%), and peer feedback mechanisms (19%).

Table 17

Ranks and Frequencies for Facilitators of Best Practices (N=289)

Rank	Facilitator	Percentage
1	Faculty development regarding exam item best practices	59.5
2	Exam item mentoring for new faculty	46.7
3	Faculty within the department with expertise in item creation, exam evaluation, and item revision	37.7
4	Testing policy which includes exam creation best practices	32.5
5	Administration support for use of exam item best practices	26.3
6	Administrative support for exam creation and revision	24.9
7	Department item developer position to support faculty	23.5
8	Available faculty development funds	19.7
9	Peer feedback mechanisms	19.0

Participants were given an opportunity to type in a facilitator if they felt something was not included in the survey question. Twenty-four participants chose to enter text and three themes were identified. As seen in Table 18, 33% percent of the participants contributed comments with the theme of additional time or workload credit to allow for the implementation of exam item best practices. Twenty-nine percent of the participants identified the use of software programs as a facilitator of exam best practices. And, one fourth of the participants listed a need for further faculty development for individuals, teams, or department experts.

Table 18

Additional Facilitators of Implementation of Best Practices

Rank	Percentage	Theme	Examples of comments
1	33%	Workload/time	Workload credit or time to develop and create exams and evaluate exams
			Give course credit to an exam champion
			More faculty time to create blueprints and test items
			Workload allocation for assessment development and evaluation
2	29%	Software	Software (like exam soft) that quickly provides test analysis
			Software limitations which prevent use of best practices
			Process in place for creation, protection, and maintenance of a central faculty-built pool of questions
3	25%	Development	Faculty will need updating
			Faculty development regarding interpreting exam results
			Continuing education in test construction and item analysis.

Research Question Six

The focus of research question six is on the associations between faculty-identified barriers and facilitators of exam item best practice implementation and faculty's demographic, individual, environmental, and social factors.

Barriers. Chi-square analyses were used to examine the relationships between the chosen barriers and the faculty's demographic, individual, environmental, and social factors. For the sake of space and clarity, only significant results are presented. Also, a table summarizing percentages and specific chi-square results is provided in each section.

Frequency of exam writing. Nursing faculty who wrote exam items less frequently were significantly more likely to select the following as barriers compared to faculty who wrote exams more frequently: "Lack of faculty development regarding item creation," "Lack of faculty development regarding exam post-test review," "Lack of faculty development regarding revision," "Lack of adequate exam item tracking software," and "Lack of testing policy that supports exam creation best practices." For most of these results, nursing faculty who wrote items "monthly" or "weekly" responded similarly to one another.

Table 19

Significant Barrier Results for Frequency of Exam Writing

Barrier	One to Two Times per Year	Two to Three Times a Semester	Monthly	Weekly	Chi-square value	Sig.
Lack of faculty development regarding item creation	74.4%	69.6%	50%	59.1%	11.34	.01
Lack of faculty development regarding exam post-test review	71.1%	69.6%	51.2%	50.5%	10.92	.012
Lack of faculty development regarding revision	76.9%	70.7%	54.1%	56.8%	9.96	.019
Lack of adequate exam item tracking software	50%	43.5%	27.8%	47.7%	9.83	.020
Lack of testing policy that supports exam creation best practices	71.1%	53.8%	45.4%	47.7%	7.87	.049

df = 3, *N* =281

Peer Mentor.   Nursing faculty who did not have a mentor (61.2%) were significantly more likely to choose "Lack of testing policy that supports exam creation best practices" as a barrier compared to those who had a peer mentor (38.2%).

Table 20

Significant Barrier Results for Peer Mentor

Barrier	Formal or informal mentor	No mentor	Chi-square value	Sig.
Lack of testing policy that supports exam creation best practices	38.2%	61.2%	6.40	.011

df = 1, *N* = 122

Years Since Completing Highest Degree.   Nursing faculty who more recently completed their highest degree (0-2 years and 3-5 years) were significantly more likely to choose "Lack of faculty development regarding revision," "Lack of administration support," and "Lack of testing policy that supports exam creation best practices" as barriers compared to those who completed their highest degree a while ago (6-10 years and More than 10 years ago).

Table 21

Significant Barrier Results for Years Since Completing Highest Degree

Barrier	0-2 Years	3-5 Years	6-10 Years	More than 10 years	Chi-square value	Sig.
Lack of faculty development regarding revision	75%	71.6%	57%	55.2%	8.60	.035
Lack of administration support	61.4%	62.5%	44.9%	41.4%	10.10	.018
Lack of testing policy that supports exam creation best practices	68.2%	58.3%	46.2%	43.7%	9.25	.026

df = 3, *N* = 281

Certified Nurse Educator (CNE).   Compared to nursing faculty who are CNEs, nurses who were not CNEs were significantly more likely to select the following as barriers: "Lack of faculty development regarding item creation," "Lack of faculty development regarding exam post-test review," "Lack of faculty development regarding revision," "Lack of funding for faculty development," and "Lack of educational preparation for exam creation and evaluation."

Table 22

Significant Barrier Results for Certified Nurse Educator (CNE)

Barrier	CNE	Not a CNE	Chi-square value	Sig.
Lack of faculty development regarding item creation	45.2%	67.8%	12.70	< .001
Lack of faculty development regarding exam post-test review	44%	66.2%	11.98	.001
Lack of faculty development regarding revision	45.9%	70.4%	15.30	< .001
Lack of funding for faculty development	50%	62.9%	4.08	.043
Lack of educational preparation for exam creation and evaluation	48.8%	70.1%	11.49	.001

$df = 1$, $N = 281$

Department Testing Policy.   Nursing faculty who teach in a department that does not have a testing policy that includes best practices were significantly more likely to choose "Time constraints," "Lack of faculty development regarding item creation," and "Lack of testing policy that supports exam creation best practices" as barriers compared to those who are in departments with a testing policy.

Table 23

Significant Barrier Results for Department Testing Policy

Barrier	Testing Policy	No Testing Policy	Chi-square value	Sig.
Time constraints	74.3%	88.5%	3.84	.05
Lack of faculty development regarding item creation	45.9%	67.8%	5.23	.022
Lack of testing policy that supports exam creation best practices	22.2%	62.8%	16.71	< .001

df = 1, *N* = 122

Exam Blueprints. Nursing faculty who teach in a department that does not utilize exam blueprints were significantly more likely to choose "Time constraints," "Lack of peer support within department," and "Lack of testing policy that supports exam creation best practices" as barriers compared to those who are in departments that use exam blueprints.

Table 24

Significant Barrier Results for Departments that Utilize Exam Blueprints

Barrier	Blueprints	No Blueprints	Chi-square value	Sig.
Time constraints	73.5%	88.6%	4.26	.039
Lack of peer support within department	26.5%	50%	5.53	.019
Lack of testing policy that supports exam creation best practices	27.3%	59.6%	10.04	.002

df = 1, *N* = 122

Exam Software. Nursing faculty who teach in a department that does not utilize software to track exam statistics were significantly more likely to select the "Lack of faculty development regarding item creation, "Lack of faculty development regarding revision," "Lack of peer support within department," "Lack of administrative support," "Lack of adequate exam

item software," and "Lack of testing policy that supports exam creation best practices" as
barriers compared to those who are in departments that use exam software.

Table 25

Significant Barrier Results for Departments that Utilize Exam Software

Barrier	Exam Software	No Exam Software	Chi-square value	Sig.
Lack of faculty development regarding item creation	54.9%	73.8%	4.20	.041
Lack of faculty development regarding revision	61%	83.3%	6.44	.011
Lack of peer support within department	36.6%	57.5%	4.79	.029
Lack of administrative support	56.8%	78%	5.34	.021
Lack of adequate exam item software	11.1%	82.9%	61.51	< .001
Lack of testing policy that supports exam creation best practices	40.7%	70.7%	9.80	.002

$df = 1, N = 122$

Administrative Support for Exams. Nursing faculty in programs that do not have
administrative support for exam creation and item tracking were significantly more likely to
select "Time constraints," "Lack of administrative support," and "Lack of testing policy that
supports exam creation best practices" as barriers compared to nurses in programs with
administrative support.

Table 26

Significant Barrier Results for Administrative Support for Exams

Barrier	Administrative Support	No Administrative Support	Chi-square value	Sig.
Time constraints	66.7%	87.4%	4.97	.026
Lack of administrative support	42.1%	67.6%	4.52	.034
Lack of testing policy that supports exam creation best practices	22.2%	56.3%	7.13	.008

df = 1, *N* = 122

Administration Support for Exam Best Practices. Nursing faculty in programs that do not have administration support for exam best practices were significantly more likely to select "Time constraints," "Lack of peer support within department," "Lack of administrative support," "Lack of administration support," "Lack of testing policy," and "Lack of funding for faculty development" as barriers compared to nurses in programs with administration support.

Table 27

Significant Barrier Results for Administration Support for Exam Best Practices

Barrier	Administrative Support	No Administrative Support	Chi-square value	Sig.
Time constraints	76.8%	90.9%	4.60	.032
Lack of peer support within department	29.1%	55.2%	8.40	.004
Lack of administrative support	48.2%	77.3%	11.09	.001
Lack of administration support	26.8%	75.8%	29.19	< .001
Lack of testing policy	37.5%	62.1%	7.35	.007
Lack of funding for faculty development	41.8%	66.7%	7.50	.006

df = 1, *N* = 122

Experience Teaching.   The primary theme with the amount of teaching experience variable is the less experience teaching, the more likely the nursing faculty will endorse one of the barriers to best practices. Nursing faculty with more than 10 years of experience teaching were significantly less likely to endorse something as a barrier.

Table 28

Significant Barrier Results for Experience Teaching

Barrier	0-5 Years of Experience	6-10 Years of Experience	More than 10 years of Experience	Chi-square value	Sig.
Lack of faculty development regarding item creation	75.8%	71.7%	52.4%	13.52	.001
Lack of faculty development regarding exam post-test review	77%	66%	51.2%	13.56	.001
Lack of faculty development regarding revision	79%	83%	50.9%	26.59	< .001
Lack of administrative support	67.7%	73%	56%	6.16	.046
Lack of administration support	63.9%	57.7%	44%	8.26	.016
Lack of educational preparation for exam creation and evaluation	77.4%	65.4%	58.1%	7.39	.025

df = 2, *N* = 281

Age.   Similar to the years of experience results, older nursing faculty were significantly less likely to select something as a barrier compared to younger nursing faculty. The specific barriers that younger faculty were more likely to select were: "Time constraints," "Lack of

faculty development regarding revision," "Lack of administrative support," and "Lack of faculty administration support."

Table 29

Significant Barrier Results for Age

Barrier	25-44 Years Old	45-54 Years Old	Older than 54 Years	Chi-square value	Sig.
Time constraints	89.6%	90.5%	79.1%	6.46	.039
Lack of faculty development regarding revision	75%	65.8%	55.7%	7.65	.022
Lack of administrative support	70.6%	70.7%	52.5%	9.78	.008
Lack of faculty administration support	64.7%	56%	41.3%	11.05	.004

df = 2, *N* = 281

Facilitators. First, chi-square analyses were used to examine the relationships between the chosen facilitators and the faculty's demographic, individual, environmental, and social factors. Again, for the sake of space and clarity, only significant results are presented. Also, a table summarizing percentages and specific chi-square results is provided in each section.

Frequency of exam writing. Nursing faculty who more frequently write exam items were significantly more likely to select "Peer Feedback Mechanisms" as a facilitator compared to those who wrote exam items with less frequency. Specifically, those who wrote items Weekly (20.5%) or Monthly (26.6%) were more likely to choose peer feedback mechanisms as a facilitator compared to those who wrote Two to Three Times a Semester (16.1%) or One or Two Times a Semester or Never (5.1%).

An odd significant pattern emerged for the frequency of exam writing by administrative support chi-square. Nursing faculty who either rarely wrote items or often wrote items

(weekly) were more likely to select administrative support as a facilitator compared to nurses in the other two categories.

Table 30

Significant Facilitator Results for Frequency of Exam Writing

Facilitator	One to Two Times per Year	Two to Three Times a Semester	Monthly	Weekly	Chi-square value	Sig.
Peer Feedback Mechanisms	5.1%	16.1%	26.6%	20.5%	9.04	.024
Administrative Support	33.3%	16.1%	25.7%	36.4%	8.34	.04

df = 3, *N* = 285

Certified Nurse Educator (CNE. Nursing faculty who are CNE's (48.2%) were significantly less likely to choose "Faculty development regarding exam item best practices" as a facilitator compared to those who are not Certified Nurse Educators (65.5%).

Table 31

Significant Facilitator Results for Certified Nurse Educator (CNE)

Facilitator	CNE	Not a CNE	Chi-square value	Sig.
Faculty development regarding exam item best practices	48.2%	65.5%	7.43	.006

df = 1, *N* = 285

Testing Policy. Nursing faculty who teach in a program with a testing policy (29.7%) were significantly more likely to choose "Peer Feedback Mechanisms" as a facilitator compared to those who are not in a program with a testing policy (13.8%), χ^2 (1, N = 124) = 4.36, p = .037.

Table 32

Significant Facilitator Results for Testing Policy

Facilitator	Testing Policy	No Testing Policy	Chi-square value	Sig.
Peer feedback mechanisms	29.7%	13.8%	4.36	.037

df = 1, *N* = 124

Exam Blueprints. Nursing faculty who use testing blueprints were significantly more likely to choose "Peer Feedback Mechanisms" and "Available faculty development funds" as facilitators compared to those who do not use testing blueprints.

Table 33

Significant Facilitator Results for Exam Blueprints

Facilitator	Exam Blueprints	No Exam Blueprints	Chi-square value	Sig.
Peer feedback mechanisms	37.1%	11.2%	11.16	.001
Available faculty development funds	34.3%	12.4%	7.99	.005

df = 1, *N* = 124

Tracking Software. Nursing faculty in departments that use software for tracking exam items (24.4%) were significantly more likely to say that availability of faculty funds is a facilitator compared to those in departments who do not use exam software (7.1%).

Table 34

Significant Facilitator Results for Tracking Software

Facilitator	Tracking Software	No Tracking Software	Chi-square value	Sig.
Available faculty development funds	24.4%	7.1%	5.47	.019

df = 1, *N* = 124

Administrative Support.　Nursing faculty in programs with administrative support for exam writing (42.1%) were significantly more likely to select "Available faculty development funds" as a facilitator compared to those who do not have administrative support (13.5%).

Table 35

Significant Facilitator Results for Administrative Support

Facilitator	Administrative Support	No Administrative Support	Chi-square value	Sig.
Available faculty development funds	42.1%	13.5%	8.97	.003

df = 1, *N* = 124

Administration Support.　Nursing faculty who say they do not have support from their administration leadership for exam writing were significantly more likely to select "Available faculty development funds" and "Administrative support for exam creation and revision" as facilitators compared to those who say they have leadership support for exam writing.

Table 36

Significant Facilitator Results for Administration Support

Facilitator	Administration Support	No Administration Support	Chi-square value	Sig.
Available faculty development funds	7%	28.4%	9.28	.002
Administrative support for exam creation and revision	12.3%	34.3%	8.16	.004

df = 1, *N* = 124

When evaluating the distribution of barriers with the faculty variables, patterns emerged. Table 37 has the frequency distribution for faculty with the same variable and the identified barriers. Table 38 has the frequency distribution for faculty with the same variable and the identified facilitators.

Table 37

Comparison Frequency of Faculty Variables and Barriers

Faculty variables		Significant Barriers										
		TC	FD_IC	FD-PTR	FD-IR	Peer	Admin	Leader	Software	Policy	FD-Fund	Edprep
Demographic	Age	X^{12}			X^{12}		X^{12}	X^{12}				
Individual	Highest level of education											
	Years since degree completion				X^{4}			X^{4}		X^{4}		
	Years of experience		X^{11}	X^{11}	X^{11}		X^{11}	X^{11}				X^{11}
	Frequency writing exam items		X^{2}	X^{2}	X^{2}					X^{2}	X^{2}	
	CNE		X^{5}	X^{5}	X^{5}						X^{5}	X^{5}
Environmental	Type of institution											
	Program has a testing policy	X^{6}	X^{6}							X^{6}		
	Administrative support	X^{9}					X^{9}			X^{9}		
	Exam software		X^{8}		X^{8}	X^{8}	X^{8}			X^{8}	X^{8}	
	Funds for faculty development											
	Utilize exam blueprints	X^{7}					X^{7}			X^{7}		
	Administration support	X^{10}				X^{10}	X^{10}	X^{10}		X^{10}	X^{10}	
Social	Faculty mentor									X^{3}		

X = likely to choose

2 chosen by faculty who wrote exams less frequently

3 no mentor

4 Faculty who recently completed highest degree
5 Chosen by faculty not CNE certified
6 chosen by faculty who do not have a testing policy
7 Chosen by faculty who do not utilize blueprints
8 Chosen by faculty who do not have exam software
9 do not have admin support
10 do not have leader support
11 less experience teaching
12 younger faculty more likely to select

Table 38

Comparison Frequency of Faculty Variables and Facilitators

Faculty variables		Significant Facilitators		
		+ FD fund	+ peer FB	+ admin.
Demographic	Age			
Individual	Highest level of education			
	Years since degree completion			
	Years of experience			
	Frequency writing exam items		X*	X**
	CNE		-	
Environmental	Type of institution			
	Program has a testing policy		X	
	Administrative support	X		
	Exam software	X		
	Funds for faculty development			
	Utilize exam blueprints	X	X	
	Administration support	X[1]		X[1]
Social	Faculty mentor			

X = likely to choose
- = less likely to choose
*Chosen by faculty who write exam questions weekly,
**admin support by weekly & those who write rarely
1 Chosen by faculty who do *not* have administration support

Further analyses

Comparing barriers and facilitators. When looking at the combined results of barriers

and facilitators, faculty with fewer opportunities to write exam items (one-two times/year and

two-three times/semester) selected the three faculty development barriers related to item

creation, posttest review, and item revision along with the lack of exam software and a testing policy that guides exam item best practices. Faculty who write exam items frequently (monthly or weekly) chose the facilitator of peer feedback mechanisms. Interestingly, the faculty who wrote exam items the least (one-two times/year) and those who wrote exam items the most (weekly) identified the facilitator of administrative support more frequently. Administrative support is regarded as a facilitator by both the ends of the exam item frequency spectrum.

Faculty who were not certified as CNE more frequently chose barriers related to development and education than faculty with CNE certification. All three barriers related to faculty development regarding, item creation, posttest review and item revision were chosen more frequently, as well as a lack of educational preparation. Relatedly, non-certified CNE faculty also chose a lack of faculty development funding. In regards to facilitators, faculty with CNE certification are less likely to choose peer feedback mechanisms.

The faculty who did not have a testing policy that guided best practices identified the barriers of time constraints, lack of faculty development regarding item creation, and a lack of a testing policy that supports exam item best practices. Faculty who have a testing policy more frequently chose the facilitator of peer feedback mechanisms.

Faculty without administrative support for exam item best practices more frequently chose the barriers of time constraints, lack of administration/leader support, and lack of testing policy than faculty who identified as having administrative support. Faculty who had administrative support selected available faculty development funds as a facilitator.

Those faculty who did not have exam software identified more barriers than faculty with exam software. These faculty also choose the facilitator of available faculty development funds. Faculty without exam software desired more faculty development opportunities and funding to

do so and identified a lack of support from peers and administrative staff as a barrier. Faculty who utilized exam software identified funds for faculty development as a facilitator of best practices.

Faculty who do not utilize exam blueprints chose the barriers of time constraints, lack of peer support, and lack of testing policy more frequently than those who utilized exam blueprints. Those who utilized exam blueprints identified available faculty development funds and peer feedback mechanisms as facilitators of exam item best practices.

Faculty without administration support for exam item best practices more frequently chose the barriers of time constraints, lack of peer, administrative, and administration/leader support, lack of exam software, and lack of testing policy than faculty who identified as having administration/leader support. Faculty without administration support also selected available faculty development funds and administration support as facilitators of exam item best practices.

Observations of related faculty variables and the barriers and facilitators. Younger faculty, those who have more recently completed their degree, have fewer years of teaching experience, and write exam items less frequently identified similar type barriers. These barriers included the lack of faculty development regarding item creation, faculty development regarding posttest review, faculty development regarding revision, administrative support, and administration/leader support.

The highest-ranking barrier of time constraints were identified more frequently by faculty with the variables of younger age, no testing policy, no administrative support, no administration/leader support, and not utilizing exam blueprints. The addition of administration and administrative support along with the structure provided by a testing policy that supports exam item best practices and the use of blueprints may cohesively address this barrier.

The highest-ranking facilitator of exam item best practices was faculty development. As mentioned earlier, the faculty variables of younger faculty, the recent completion of their highest degree, fewer years of experience teaching, and those who wrote exam items less frequently indicated the lack of faculty development as a barrier. Interestingly, faculty who were not CNE also were more likely to choose the barriers of lack of faculty development.

Administrative support was a barrier associated more frequently with the faculty variables of younger age, fewer years since highest degree completion, fewer years of experience, and those without administrative support, administration support, and exam software.

Another barrier that was frequently associated with numerous faculty variables was the lack of a testing policy that supports exam item best practices. The faculty variables included faculty with fewer years since highest degree completion, who wrote exam items less frequently, who did not have a testing policy that included exam item best practices, those who utilized blueprints, and those without administrative support, administration support, and exam software.

The facilitator of available funds for faculty development was significantly associated with the faculty variables of faculty who do utilize exam blueprints and those without administrative support, administration support, and exam software.

Conclusion

The results of the survey identified faculty variables that predicted the use of exam item best practices among the demographic and individual variable groups. Faculty ranked barriers to and facilitators of exam item best practices. Associations between faculty variables and barriers to and facilitators of exam item best practices were also identified. The following chapter will include a discussion regarding these findings and explore potential implications.

Chapter 5: Discussion, Implications, and Recommendations

Overview of Study

The first purpose of this study was to examine relationships between nursing faculty's demographic, individual, environmental, and social factors and their knowledge and use of exam item best practices. The second purpose was to examine what nursing faculty identified as potential barriers and facilitators of exam item best practices implementation. In addition to indicating which exam item creation best practices faculty regularly utilize, participants identified which facilitators and barriers for the implementation of exam item best practices they consider most prevalent in their setting. These data could prove instrumental in the development of nursing departments' delegation of resources, policies regarding exam item best practices, the use of blueprints, establishing peer feedback mechanisms, and other facilitators of exam best practices support (Birkhead et al., 2018; Halstead, 2013; Ibrahim, 2019; Lavin & Rosario-Sim, 2013; Obon & Rey, 2019; Ray et al., 2018; Tarrant & Ware, 2012). Ultimately, the implementation of better exam items improves the preparation of graduates, which leads to better, safer, more efficient patient care (Betts et al., 2019; Dickerson et al., 2019; Oermann & Gaberson, 2021).

Synthesis of the Results

Individual variables. The three individual factors of years of experience, frequency of writing exam items, and CNE certification were found to be predictors of a higher exam best practice (EBP) score. This aligns with what other researchers have found. Moore (2020) found that faculty with more than 5 years of experience tended to use best practices for exam creation, analysis, and revision. This also aligns with Benner's (1984) theory of novice to expert. The number of years of experience and frequency of writing exam items correlate with the

demographic variable of faculty age. The longer faculty practice and the more exam items they create would allow for greater opportunities to improve their ability to write well-developed exam items. The data also showed that younger faculty newer to nursing education and with less experience writing exam items identified more barriers than older faculty with more experience teaching and with item writing. Faculty who identified more barriers wanted support in the form of education and development, policies that embedded best practices, exam software, and support from administrative personnel and administration (leadership).

Faculty who obtained CNE certification also attained a higher EBP score.   Part of the CNE certification includes a section entitled *Use Assessment and Evaluation Strategies* (National League for Nursing, 2019). In this section, exam best practices are embedded as expectations of assessment practice. As faculty prepare for the certification exam, they are exposed to exam item best practices. This exposure provides faculty with the opportunity to review exam item best practices and be better equipped to implement them in their setting. Those faculty who were not CNE prepared identified more barriers related to educational preparation, development, and a lack of faculty development funding. A faculty with more CNE certified members may require less faculty development regarding exam item best practices and free up time and resources to address other areas in nursing education.

Barriers. The highest-rated barrier of the 11 barriers identified in the literature was time constraints, which aligns with recent literature (Betts et al., 2019; Birkhead et al., 2018; Ibrahim, 2019; Khafagy et al., 2016; Ray et al., 2018; Rudolph et al., 2019). The barriers associated with faculty development were rated among the top barriers, with lack of faculty development regarding item revision, lack of faculty development regarding item creation, and lack of faculty development regarding posttest review ranked third, fifth, and sixth respectively. Combined

with the second-ranked barrier of lack of educational preparation, faculty clearly indicated that a lack of education and faculty development presented a barrier to the implementation of exam item best practices. To further support this conclusion, 35% of faculty's comments in the optional text entry question regarding barriers contained the theme of lack of educational preparation and/or faculty development.

Faculty ranked lack of administrative support as fourth. Examples of administrative support on the survey included: administrative personnel who enter exam items into LMS or track item statistics. This ranking links directly with the time constraint barrier identified by faculty as number one. With the support of administrative personnel to tend to these types of tasks, faculty may have time to focus on item creation, analysis, and revision.

Although the availability of faculty development funds ranked eighth, this facilitator fits into the faculty development category. Those who desired more faculty development also identified funding as a significant facilitator. Faculty ranked lack of administration support as ninth. Lack of administration support for exam item best practices was defined for the participants as: does not advocate for faculty knowledge development, time for exam item development, and funding to support this work. This result directly correlates with other barriers such as time constraints and lack of faculty development.

Lack of peer support ranked tenth; however, 30% of the participants who chose to write in text for items not mentioned in the Qualtrics matrix regarding barriers identified a theme of lack of peer/colleague support. Similarly, lack of exam software ranked eleventh with 10% of participants identifying LMS barriers in the optional text entry question for barriers.

Facilitators. Faculty were asked to choose what they felt were the top three facilitators from a list of nine facilitators developed from the literature. The top facilitator chosen by faculty

was faculty development regarding exam item best practices. Given high rankings in the barrier section associated with lack of faculty development and lack of educational preparation, this result is expected.   The second-ranked facilitator of exam item best practices was mentoring for new faculty, which also relates to the faculty's perception of how significant development and expert support are in the use of exam item best practices.   Although mentorship is strongly suggested in the literature as a key component to exam best practices implementation (Birkhead et al., 2018; Halstead 2013; Ibrahim, 2019; Lavin & Rosario-Sim, 2013; Obon & Rey, 2019; Ray et al., 2018; Rudolph et al., 2019; Tarrant & Ware, 2012), faculty who identified that they had a formal or informal mentor did not achieve a higher EBP score than those who did not have a mentor. Mentorship was not narrowly defined in the survey as a mentor who assists with exam item creation, analysis, and revision; therefore, faculty who had a mentor did not necessarily receive mentor support for exam item best practices.   It is noteworthy that participants recognized a mentor specific to exam item development would be a desirable facilitator.

The third facilitator identified by faculty was faculty within the department with expertise in item creation, exam evaluation, and item revision. The theme of faculty support to better implement exam item best practices continued and one could argue that the faculty expert could easily act as the mentor for new faculty, making the second and third-ranked facilitators closely related and significant.   Another means of supporting the use of exam item best practices is to have a testing policy which includes exam creation best practice (Barton et al., 2014; Oermann & Gaberson, 2021), This facilitator was ranked fourth. A department item developer position to support faculty also falls into this umbrella of supporting the use of item best practices. This facilitator was ranked seventh.

Faculty were given the opportunity to enter facilitators in text form if they felt a facilitator was not represented in the list. The theme of compensating for workload and time was mentioned by 33% of the participants. Some specific examples included giving workload credit or creating an "exam champion" to offset the time constraints of exam best practices use. Another theme identified in the text option was software to assist with item development, test analysis, and exam security. The final theme returned to the topic of needed faculty development.

When evaluating what faculty identified as top barriers and facilitators, the theme of faculty support via faculty development and expert mentorship emerges. This is supported in the recent literature (Betts et al., 2019; Cox, 2019; O'Rae et al., 2019; Oermann & Gaberson, 2021; Rudolph et al., 2019). With improved preparation and support for the use of exam item best practices, the researcher speculates that the time constraints would decrease as faculty became more proficient at item creation, analysis, and revision.

Implications for Practice/Leadership/Nursing Programs

Faculty variables. Faculty variables that predicted a higher exam best practice score included older faculty, those with more experience teaching, increased frequency of exam item writing, and certification as a certified nurse educator (CNE). Nursing leadership should acknowledge that experience matters and create opportunities for faculty with more expertise to mentor newer and less experienced faculty. To collectively improve faculty ability to create and maintain high-quality exams, time for writing and analyzing exam items should be regularly set. These opportunities will provide feedback and mentoring for faculty (Khafagy et al., 2016), as well as better assessments and formation of students' clinical judgment skills (e.g. Betts et al., 2019; Oermann & Gaberson, 2021; Rudolph et al., 2019). Nursing leaders should cultivate the

expectation that faculty will participate in giving and receiving feedback in order to improve exam items. The importance of developing a student's clinical judgment should act as the impetus for support.

Given the higher scores for those who have obtained their CNE, nursing leaders should encourage faculty to obtain this certification and provide financial support for its completion. Considering that those with CNE certification were less likely to choose the facilitator of faculty development regarding exam item best practices, these faculty may be the experts within departments who provide mentorship and support for newer, less experienced faculty.

Faculty Development. The data strongly indicated that nursing leaders should consider faculty development regarding exam item creation, analysis, and revision. This skill will become more crucial as the National Certification for Licensure Exam (NCLEX) adopts the NextGen format questions and continues to use multiple-choice questions to assess higher-order thinking and the ability of the graduate to exercise clinical judgment (e.g., Bett et al., 2019, Dickison et al., 2019). The preparation of nursing educators should also be evaluated as this barrier was associated with faculty with fewer years of experience and who are not CNE. With the movement away from the master's degree in nursing education, fewer faculty will receive content regarding exam item best practices prior to beginning an academic career and increase the need for faculty development in this area.

When evaluating the focus of faculty development, the area identified most consistently was faculty development regarding item revision. Item revision improves item and exam performance (D' La & Visbal-Dionaldo, 2017; Khafagy et al., 2016; Obon & Rey, 2019; Oermann & Gaberson, 2021; Rudolph et al., 2019). Of the best practices included in the posttest review matrix on the Qualtrics survey, track distractor efficacy scored the lowest (2.02, N =

285). On the survey's item revision matrix, "use statistical data to revise item distractors" scored lower than "'use statistical data to revise item stems" with an average score of 2.34 compared to 2.54 (N = 285). The combination of these two results highlights the underutilized best practice of item distractor tracking and subsequent revisions to strengthen exam items. Although writing strong stems sets the stage for an excellent item, the distractors determine the item's discriminatory ability (Khafagy et al., 2016).

Mentorship. The faculty variable of mentorship was not a predictor of a higher EBP score; however, mentorship specific to exam items was ranked second of the facilitators identified by faculty. Beyond the possibility of using faculty with CNE, mentorship specific to exam item best practices and/or a faculty expert who works with faculty to assist with exam best practices implementation should be evaluated as a support for newer, less experienced faculty. An alternative solution would be to pair faculty who write exam items more frequently with the new, less experienced faculty. Pairing newer faculty, who have been teaching for less than five years, with faculty who have more than five years of teaching experience with exam item creation, analysis, and revision can best support faculty development in this area and improve the program's assessments (Moore, 2020). The expectation that newer faculty are expected to consult with the designated mentor should also be clear. Nursing leaders may consider how to best utilize the faculty who have earned CNE in a mentorship role as well.

Exam software. Although the lack of exam software ranked eighth on the barriers list, 66% (N = 124) of participants indicated that they use exam software (e.g., ExamSoft®). Within this context, the lower ranking for this barrier is better understood. Of the faculty who did not have exam software available to them, more barriers to the implementation of exam best practices were identified. The barriers more commonly identified were: "lack of faculty

development regarding item creation," "lack of faculty development regarding item revision,"

"lack of peer support," "lack of administrative support," "lack of exam software," and "lack of

testing policy to support exam item best practices." Given the emphasis on tracking and

statistical evaluation, the addition of exam software can further augment faculty's ability to use

exam item best practices (D' La & Visbal-Dionaldo, 2017; Khafagy et al., 2016; Obon & Rey,

2019; Oermann & Gaberson, 2021; Rudolph et al., 2019). Of note, faculty without exam

software identified the need for faculty development funds more frequently. The use of exam

software provides faculty with the ability to more easily build exams to meet the requirements of

blueprints, statistically evaluate items and exams, and track the performance of items after

revision. Nursing departments that do not currently use exam software should evaluate the

cost/benefit of investing in this tool for improved assessments.

       Administration/Leader support. Among faculty who did not have administration or

leaders' support, numerous barriers were identified including time constraints, lack of peer

support, lack of administrative support, lack of testing policy, and lack of faculty development

funding. Leaders directly influence the learning environment and can facilitate faculty

development, evaluate workloads, designate resources, and create and support policies with clear

guidelines (Josiah Macy Foundation Jr., 2018; Khafagy et al., 2016; Morrill, 2010; Witherspoon,

1997). For younger, less experienced faculty, administration support was identified more

frequently. Nursing leaders can address this finding by assuring that new faculty members have

a mentor specific to exam item writing and ensure testing policies, including the use of

blueprints, are set as clear guidelines (Eweda et al., 2020). Since nursing leaders also can impact

the distribution of resources, several of the barriers and facilitators can be addressed.

Availability of faculty development funds emerged as a significant facilitator for faculty without

administrative and administration support, without exam software, and for those with testing policies. Beyond providing funds for faculty development, obtaining faculty development, the use of exam software, and the level of administrative assistance available to faculty are potential areas where the allocation of resources could support the implementation of exam best practices.

Administrative support. In conjunction with administration or leadership support, administrative support can impact faculty's ability to implement exam item best practices. Of note, faculty identified not having administrative support as fourth on the barriers list, and sixth on the facilitators list. Both young, newer faculty and those who write exam items most frequently identified the lack of administrative support as a significant barrier. Off-loading some tasks to administrative support staff will increase the amount of time that faculty have to create, analyze, and revise exam items.

Testing policy. As previously mentioned, the addition of administration and administrative support along with the structure provided by a testing policy that supports exam item best practices and the use of blueprints may cohesively address the barrier of time constraints. With a testing policy that supports exam best practices in place, improved assessment practices with better graduate clinical judgment development also are possible. A testing policy that supports exam item best practices includes the utilization of blueprints. Blueprints allow faculty to design exams that ensure testing of intended material at the desired cognitive level (Eweda et al., 2020; Oermann & Gaberson, 2021). Testing policies can also guide the scaffolding of higher-order thinking questions to develop clinical judgment across the curriculum (Betts et al., 2019; Oermann & Gaberson, 2021; Scully, 2017).

With the testing policy, requirements for peer feedback should be embedded. Although peer feedback did not rank high on the barriers or facilitators list, the data indicated that those

faculty who write exam items frequently, faculty who have a testing policy, and faculty who utilized blueprints identified peer feedback mechanisms as a facilitator more frequently than those without these faculty variables. Peer feedback may occur within teaching teams or may be offered by someone who does not teach within the course. Similar to the development of clinical judgment, faculty development of writing exam items can be enhanced by peer discussion and feedback and serve in the dual roles of quality assurance and mentoring. The peer feedback mechanisms should occur during the multiple phases of exam creation, analysis, and revision. This approach supports Benner's (1994) novice to expert theory and is recommended in recent literature (Betts et al., 2019; Dickinson et al., 2019; Eweda et al., 2020; Oermann & Gaberson, 2021).

The flowchart Implementation of Exam Item Best Practices (Figure 2) summarizes the necessary elements needed within departments to successfully implement exam items best practices. The two main branches are faculty knowledge and support. Each branch contains subcategories discussed in this chapter. With these elements in place, departments will have the ability to implement exam item best practices and improve their multiple-choice assessments.
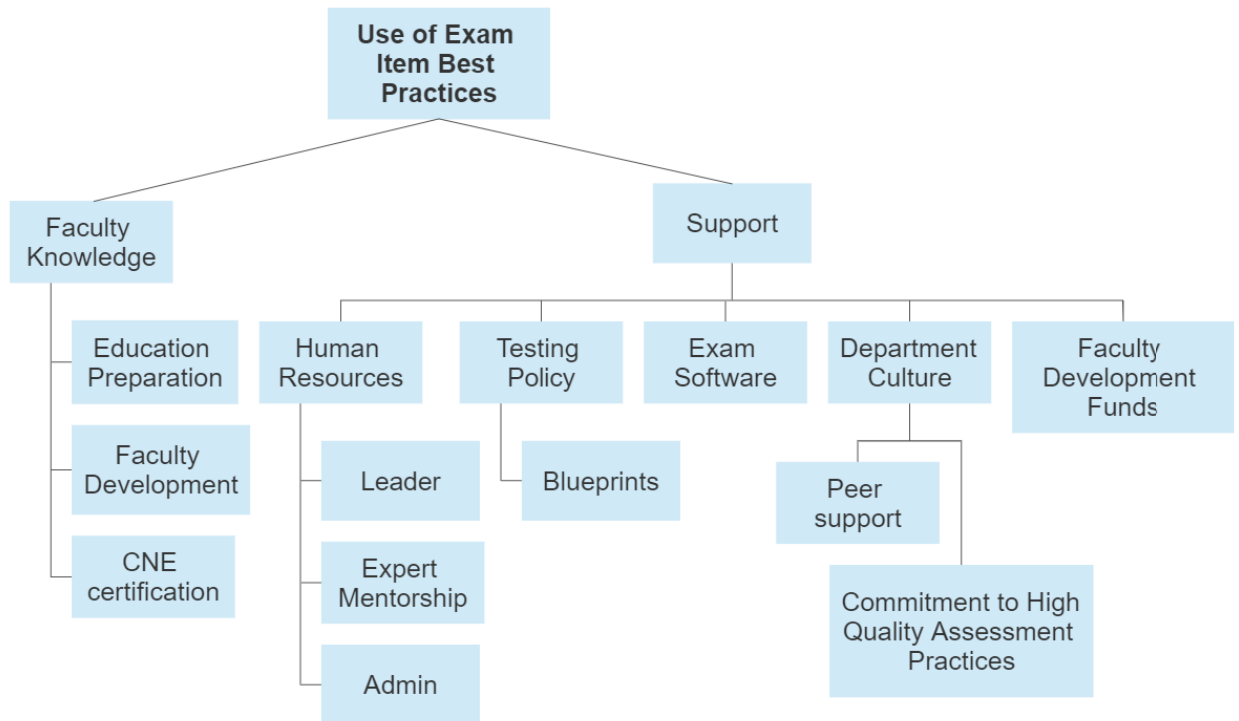
Figure 2. Implementation of Exam Item Best Practices

Limitations

    Sample Size. The number of participants may have been influenced by the gatekeeper method used to disseminate the survey instrument. Estimation of response rate and true faculty population were both difficult to calculate. As a result, the participant pool did not represent a truly random sample of nurse educators in higher education. Since participants were allowed to skip questions if desired, the sample size for questions varied.

    Length of the survey. As seen in the completion rate of questions, fewer participants answered items as the survey progressed. The time estimate of 15 minutes exceeded the recommended time limit of seven minutes. The survey contained 60 matrix style items in Qualtrics, which also exceeded the recommended number of 20.   Many participants did not complete the last page of the survey, as evidenced by the sharp decline in response rates between

the CNE question (N = 285) on the bottom of the second to the last page to the question

regarding testing policies (N = 123) on the top of the last page.

Interest. Nursing faculty who held a greater interest or investment in exam item

creation may have been more inclined to answer the survey (Orcher, 2014). Faculty who feel ill-

equipped in the area of exam item creation or do not regularly write exam items in their practice

may choose not to participate in the survey. Either scenario may be seen as a limitation.

Social influences. The survey was distributed in the spring of 2021 during the

coronavirus pandemic. The level of faculty fatigue secondary to teaching during this time frame

may have decreased the amount of time faculty were willing to take to complete surveys

(McMurtrie, 2020).

Recommendations for Future Research

This research focused on faculty use of best practices and the barriers to and facilitators

of the implementation of exam item best practices. Further areas of research should be

conducted regarding the consistent use of exam item best practices and their relationship to the

development of student clinical judgment and NCLEX pass rates. The consistent use of exam

item best practices not only includes the best practices but also should focus on the scaffolding of

assessment to move the learner from novice to advanced beginner. As a result, the research

study would investigate the assessment practices across the curriculum.

Additional research should be done to identify which faculty identified barriers and

facilitators of exam best practices implementation would be most beneficial to address,

remaining cognizant of faculty time and available resources. Since this research pointed toward

the need for a robust support system, nursing leadership may desire further research to identify

the best starting point.

Research regarding how faculty development addressing the least utilized exam best practices could impact overall use of exam best practices and improvement of exam quality and clinical judgment development. The data identified an underutilization of distractor development and revision of distractors based on post hoc analysis. Addressing this area of exam best practices may prove to be an effective way to improve multiple-choice tests.

A comparison of the type of educational preparation faculty members obtain prior to entry into academia with how comfortable they are with exam item creation, analysis, and revision is another recommendation for a future study. Since nursing faculty may have a variety of higher education degrees, different levels of preparation for exam writing exist (Obon & Rey, 2019; Tarrant & Ware, 2012). The current movement away from a Master's degree in nursing education may directly impact the preparation of nursing faculty to use exam item best practices. A study of this type could identify gaps in educational preparation regarding assessment practices, which include exam item best practices, and drive the development of new faculty orientation programs.

Concluding Comments

The ultimate goal of nursing education is to prepare our graduates to provide excellent, safe patient care. The implementation of exam item best practices can significantly contribute to this process. There continues to be an inconsistent use of exam item best practices. Faculty participants in this research identified ways in which the implementation of best practices could be supported. Nursing leadership needs to begin with an evaluation of faculty educational preparation for assessment practices which includes exam item best practices. Faculty development should continue, especially during the early years of teaching experience, but also with changes and updates to best practices and national exam standards. Designated funds will

be required to best support faculty development. One effective way to encourage faculty development is to encourage faculty to earn Certified Nurse Educator (CNE) accreditation.

Beyond education and faculty development, nursing departments should create a supportive environment for the use of exam item best practices. This support system includes a testing policy with embedded exam item best practices that require the use of blueprints and incorporates feedback mechanisms, exam software, and mentorship regarding exam item best practices for newer faculty, administrative, and administration support. This support system requires adequate resources and committed leadership. With faculty development and the support system in place, the top-ranked barrier of time constraints will be addressed and faculty more equipped to implement exam item best practices.

References

American Association of Colleges of Nursing. (2021a). *AACN Member Schools*. Retrieved

from: https://www.aacnnursing.org/About-AACN/Who-We-Are/Member-Schools

American Association of Colleges of Nursing. (2021b). *CCNE Accreditation*. Retrieved from:

https://www.aacnnursing.org/CCNE

Barton, L., Wilson, P., Langford, R., & Schreiner, B. (2014). Standardized predictive testing:

Practices, policies and outcomes. *Administrative Issues Journal, 4*(2), 68-76.

doi:10.5929/2014.4.2.2.

Bearman, M., Dawson, D., Hall, M., Bennet, S., Molloy, E., & Joughlin, G. (2014a).

Assessment Design Decisions Framework. Retrieved from Assessment Design

Decisions: http://www.assessmentdecisions.org/wp-content/uploads/2014/09/Appendix-

A-the-Assessment-Design-Decisions-Framework.pdf

Bearman, M., Dawson, P., Boud, D., Hall, M., Bennett, S., Molloy, E., & Joughin, G. (2014b).

Guide to the Assessment Design Decisions Framework. Retrieved from:

http://www.assessmentdecisions.org/wp-content/uploads/2014/09/Guide-to-the-

Assessment-Design-Decisions-Framework.pdf

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for

assessment practice: Developing the Assessment Design Decisions Framework.

Teaching in Higher Education, 21(5), 545-556.

Benner, P. (1984). *From novice to expert: Excellence and power in clinical nursing practice.*

Menlo Park, CA: Addison-Wesley Pub. Co., Nursing Division.

Betts, J., Muntean, W., Kim, D., Jorion, N., & Dickison, P. (2019). Building a method for writing clinical judgment items for entry level nursing exams. *Journal of Applied Technology, 20*(S2), 21-36.

Birkhead, S., Kelman, G., Zittel, B., & Jatulis, L. (2018). The prevalence of multiple-choice testing in registered nurse licensure-qualifying nursing education programs in New York state. *Nursing Education Perspectives, 39*(3), 139-144.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals.* New York, NY: Longmas, Green.

Bogdan, R., & Biklen, S. (2007). *Qualitative Research for Education: An introduction to theories and methods,* (5th ed.). Boston, MA: Pearson Education, Inc.

Boud, D. (1995). Assessment and learning: Contradictory or complementary. In P. Knight (Ed.), *Assessment for Learning In Higher Education* (pp. 35-48). London, United Kingdom: Kogan Page.

Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment and Evaluation in Higher Education, 41*(3), 400-413. https://doi.org/10.1080/02602938.2015.1018133.

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

Coffman, J., Rencis, J. J., Jensen, D. J., Brown, A. O., Liu, J., Kaufman, K., & White, C. (2010). Development and assessment of a multiple-choice quiz for a bending stress quiz. *2010 Midwest Section Conference of the American Society for Engineering Education*, (pp. 1-22).

Cox, C. W. (2019). Best practice tips for the assessment of undergraduate nursing students via multiple choice questions. *Nursing Education Perspectives, 40*(4), 228-230.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.).* Thousand Oaks, CA: Sage Publications, Inc.

D' La, J., & Visbal-Dionaldo, M. (2017). Analysis of multiple choice questions: Item difficulty, discrimination index and distractor. *International Journal of Nursing Education, 9*(3), 109-114.

Dawson, P., Bearman, M., Boud, D. J., Hall, M., Molloy, E. K., Bennett, S., & Joughin, G. (2013). Assessment might dictate the curriculum, but what dictates assessment? *Teaching & Learning Inquiry, 1*(1), 107-111.

Department of Health, Education, and Welfare. (1979). *The Belmont Report.* Retrieved from http://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/

Dickison, P., Haerling, K., & Lasater, K. (2019). Integrating the national council of state boards of nursing clinical judgment model into nursing educational frameworks. *Journal of Nursing Education, 58*(2), 72-78.

Eweda, G., Bukhary, Z. A., & Hamed, O. (2020). Quality assurance of test blueprinting. *Journal of Professional Nursing, 36*, 166-170.

Foundation, J. M. (2018). *Improving environments for learning in the health professions: Recommendations from the Macy Foundation Conference.* Retrieved from https://macyfoundation.org/assets/reports/publications/june2018_summary_webfile_7.20.18.pdf

Halstead, J. A. (2013). The NLN's fair testing imperative and implications for faculty development [President's Message]. *Nursing Education Perspectives, 34*(2), 72.

Hijji, B. & Mahmoud, B. (2017). Flaws of multiple choice questions in teacher-constructed

    nursing examinations: A pilot descriptive study. *Journal of Nursing Education, 56*(8),

    490-503.

Ibrahim, A. F. (2019). Development of multiple choice questions' instructional plan for the

    nursing educators. *Journal of Nursing Education and Practice, 9*(9), 12-24.

Khafagy, G., Ahmed, M., & Saad, N. (2016). Stepping up of MCQs' quality through a multi-

    stage reviewing process. *Education for Primary Care, 27*(4), 299-303.

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.

Laguilles, J. S., Schweiger, D. M., & Sapienza, H. J. (2011). Can lottery incentives boost web

    survey response rates? Findings from four experiments. *Research in Higher Education,*

    52(5), 537-553.

Lavin, J., & Rosario-Sim, M. (2013). Understanding the NCLEX: How to increase the success

    on the revised 2013 examination. *Nursing Education Perspectives, 14*(3), 196-198.

Libner, J., & Kubala, S. (2017). Improving program NCLEX pass rates: Strategies from one

    state board of nursing. *Nursing Education Perspectives, 38*(6), 325-329.

Mager, D., Beauvais, A., & Kazer, M. (2017). Surviving the NCLEX dip. *Nursing Education*

    Perspectives, 38(5), 283-285.

Mambwe, P. (2017). Knowledge of Bloom's hierarchy of cognitive levels of educational

    objectives and its use in setting examinations by nurse educators in Lusaka and Eastern

    provinces. *Texila International Journal of Nursing, 3*(2), 1-13.

McMurtrie, B. (2020, November 4). *The pandemic is dragging on. Professors are burning out.*

    Retrieved from: https://www.chronicle.com/article/the-pandemic-is-dragging-on-

    professors-are-burning-out

Merriam, S. B., & Tisdel, E. J. (2016). *Qualitative Research: A guide to design and*

    implementation. San Francisco, CA: Jossey-Bass.

Moore, W. (2020). Does faculty experience count? A quantitative analysis of evidence-based

    testing practices in baccalaureate nursing education. *Nursing Education Perspectives,*

    42(1), 17-21.

Morrill, R. (2010). *Strategic leadership: Integrating strategy and leadership in colleges and*

    universities. Lanham, MD: Rowman and Littlefield.

Muijs, D. (2011). *Doing quantitative research in education with SPSS 2nd edition.* Thousand

    Oaks, CA: Sage.

Naeem, N., van der Vleuten, C., & Alfaris, E. A. (2012). Faculty development on item writing

    substantially improves item quality. *Advances in Health Science Education, 17*(3), 369-

    376.

National Council State Boards of Nursing. (2018). Strategic practice analysis. *NCSBN*

    Research Brief, 71, 1-16.

National League for Nursing. (2012). *Fair Testing Guidelines.* Retrieved from:

    http://www.nln.org/docs/default-source/advocacy-public-policy/fair-testing-

    guidelines.pdf?sfvrsn=6

National League for Nursing. (2019, July 18). *Certified Nurse Educator (CNE) 2-10 Candidate Handbook.* Retrieved from: http://www.nln.org/docs/default-source/professional-development-programs/certified-nurse-educator-(cne)-examination-candidate-handbook.pdf?sfvrsn=2

Obon, A. M., & Rey, K. A. (2019). Analysis of multiple-choice questions (MCQs): Item and test statistics from the 2nd year nursing qualifying exam in a university in Cavite, Philippines. *Abstract Proceedings International Scholars Conference, 7*(1) (pp. 499-511). Cavite City, Philippines: International Scholars Conference. https://doi.org/10.35974/isc.v7i1.1128.

Oermann, M. H., & Gaberson, K. B. (2021). *Evaluation and Testing in Nursing* Education (6th ed.). New York, NY: Springer.

Omer, A. A., Abdulrahim, M. E., & Albalawi, I. A. (2016). Multiple choice questions put on the scale: What is their impact on students' achievement in a final undergraduate surgical examination? *Journal of Health Specialties, 4*(4), 270-275.

O'Rae, A., Hnatyshyn, T., Bock, A. J., Mannion, C., & Patek, S. (2019). An appreciative inquiry into nursing educators' exam practices. *Nursing Education Perspective, 40*(4), 234-235.

Orcher, L. (2014). *Conducting Research: Social and behavioral science methods* (2nd ed.). Glendale, CA: Pyrczak Publishing.

Patten, M. (2014). *Understanding Research Methods,* (9th ed.). Glendale, CA: Pyrczak Publishing.

Ray, M. E., Daugherty, K., Lebovitz, L., Rudolph, M., Shuford, V., & DiVall, M. (2018). Best

    practices on examination construction, administration, and feedback. *American Journal*

    of Pharmaceutical Education, 82(10), 1127-1132.

Roberts, C. M. (2010). *The dissertation journey: A practical and comprehensive guide to*

    planning, writing, and defending your dissertation (2nd ed). Thousand Oaks, CA:

    Corwin.

Rudolph, M. J., Daugherty, K. K., Ray, M. E., Shuford, V. P., Lebovitz, L., & DiVall, M. V.

    (2019). Best practices related to examination item construction and post-hoc review.

    American Journal of Pharmaceutical Education, 83(7), 1492-1503.

Schroeder, J. (2013). Improving NCLEX-RN pass rates by implementing a testing policy.

    Journal of Professional Nursing, 29(2S), S43-S47.

    http://dx.doi.org/10.1016/j.profnurs.2012.07.002.

Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking.

    Practical Assessment, Research, & Evaluation, 22(4), 1-13.

Tarrant, M., & Ware, J. (2012). A framework for improving the quality of multiple choice

    assessments. *Nurse Educator, 37*(3), 98-104.

White, K. A., & Heitzler, E. T. (2018). Effect of increased evaluation objectivity on grade

    inflation: Precise grading rubrics and rigorously developed tests. *Nurse Educator, 43*(2),

    73-77.

Witherspoon, P. D. (1997). *Communicating leadership: An organizational perspective.*

    Needham Heights, MA: Allyn and Bacon.

Appendix A: Cover Letter to Deans and Directors

Dear Nursing Education Leader,

With the addition of **Next Gen** questions to the NCLEX next fall, writing rigorous, high-quality exam questions becomes more important than ever. The use of exam item best practices increases the validity of exam questions in nursing education and better prepares students for taking the NCLEX and entry level practice. Are nursing faculty consistently using exam item best practices identified by education experts? What do faculty identify as facilitators and barriers of the implementation of exam item best practices?

My name is Julie De Haan and I am a doctoral student at Bethel University, St. Paul, Minnesota. For my dissertation, I am examining nursing faculty use of best practices in exam item creation, analysis and revision and implementation barriers and facilitators. As nurse administrator of an entry-level RN program, I am contacting you to ask you to pass along this survey to your faculty who may be interested in participating in this research study. You are under no obligation to share this email with your nurse educators.

*The survey will require approximately **15 minutes** to complete.* Participants may choose to be entered for one of ten $75 Amazon.com gift cards. There are no known risks in responding to the survey. Be assured that all data collected will remain **anonymous**.

Thank you for your time and assistance in this important endeavor. If you require additional information or have questions, my contact information is included below.

Sincerely,

Julie De Haan MSN, RN
Principal Investigator
Bethel University
j-de-haan@bethel.edu
612-390-8491

Appendix B: Qualtrics Survey

Exam Item Best Practices and Implementation Barriers and Facilitators

Welcome to the study

Use of Best Practices in Item Creation, Analysis, and Revision:
Nursing Faculty's Knowledge, Use, and Implementation Facilitators and Barriers

Thank you for participating in this study regarding nursing faculty's knowledge and use of exam item best practices and exam item best practices implementation barriers and facilitators. Your participation will help identify which exam item best practices nursing faculty use most frequently and what facilitators and barriers faculty identify to the implementation of exam item best practices. Exam item best practices and the facilitators and barriers to exam item best practices were identified in current and relevant literature.

The survey will take approximately 15 minutes to complete.

After completing this survey, you will have the opportunity to enter into a drawing for one of ten $75 Amazon.com gift cards.

The survey is divided into three parts:

1. Knowledge and use of exam item best practices

2. Facilitators of and barriers to the implementation of exam item best practices

3. Faculty demographics & variables

Your individual responses will remain **anonymous**.

Informed Consent

Topic and Purpose

This survey is part of a research project for a dissertation in Bethel University's Doctor of Education in Higher Education Leadership program. You will be asked questions regarding your use of exam item best practices and implementation barriers and facilitators. Exam item best practices and the facilitators and barriers to exam item best practices were identified in current and relevant literature.

 The study has two purposes:

1. To examine the relationships between nursing faculty variables (demographic, individual, environmental, and social factors) and their use of exam item best practices.
2. To examine what facilitators and barriers of exam item best practices implementation faculty identify as prevalent.

Participants

You have been invited to participate in this study because you were identified as a nursing faculty member. There will potentially be 500 participants in this study.

Voluntary Participation

Participation in this study is voluntary. You may skip any question on the survey or discontinue participation at any time.

The researcher may remove participants from the study if survey data is incomplete.

Survey Protocol

The survey will take most participants 15 minutes to complete. It contains three sections of multiple-choice questions with potential answers or answers with a Likert scale.

Confidentiality

Your identity will not be linked to your responses on the survey and therefore remain anonymous. The survey will not collect your name, email address, or exact locations. In order to sign up for the drawing for the Amazon.com gift card and/or an executive summary of the research, you may follow a link to a google form, outside of the survey, to provide your name and email address. The information on the google form will not be linked to the responses within the survey. Your identity will remain anonymous within the survey. The names and email addresses of the participants who request to participate in the drawing and/or the executive summary will be deleted once the study is complete.

Risks and Benefits

No risks to participants have been identified. Participants may skip questions or exit the survey at any time.

Participants may request an executive summary of the study findings. If you wish to receive a summary, please complete the form linked at the conclusion of the survey or by emailing j-de-haan@bethel.edu.

Participants may also elect to be entered into a drawing for one of ten $75 Amazon.com gift cards.

Future use of data

Response data collected via this survey may be used in future studies. However, participants' identity will not be linked to the response data.

This research has been approved in accordance with Bethel University's Levels of Review for Research with Humans.

You are making a decision whether or not to participate in this study. If you do not wish to participate, you may close the browser. By clicking the button below, you acknowledge:

I have reviewed the informed consent page and I agree to participate in this study. I understand that I may skip any question in the survey and I may withdraw or discontinue participation before submitting the completed survey.

 o I consent, begin the study

o I do not consent, I do not wish to participate

Section 1: Knowledge and Use of Best Practices for Exam Creation

Your honest participation will provide actionable information.
Please respond according to how you practice <u>in reality,</u> not ideally.

Best practices: Pre-Exam Planning

For the following items, please honestly rate how frequently you use the practice in your pre-exam planning. If you are unfamiliar with the practice, please choose "not familiar with the practice".

	Not familiar with practice	Do not use	Sometimes	Half of the time	Most of the time
Develop a blueprint <u>as defined below</u>; if your blueprint does not contain all the elements, please select not familiar with practice or do not use. <u>*Definition*</u>*: document containing identified course and module objectives to be tested, the distribution of item type, cognitive level, and associated nursing process steps and/or*	o	o	o	o	o

client need categories					
Define exam purpose, congruent with module/course objectives	o	o	o	o	o
Specify the number of items on the exam and the proportion of items per content area/objectives	o	o	o	o	o

Plan for a sufficient number of items (50-60 items to achieve high level of reliability)	o	o	o	o	o
Specify the desired difficulty and discrimination level of the items	o	o	o	o	o
Specify the format of exam items, including number of alternative items	o	o	o	o	o

Best practices: General for Item Creation

For the following items, please honestly rate how frequently you use the practice in your exam planning. If you are unfamiliar with the practice, please choose "not familiar with the practice".

	Not familiar with practice	Do not use	Sometimes	Half of the time	Most of the time
Write exam items free of grammatical and structural errors, good technical quality	o	o	o	o	o
Use appropriate vocabulary (avoid colloquialisms or slang terms)	o	o	o	o	o
Write questions to test higher cognitive thinking	o	o	o	o	o
Use clinical vignettes/ scenario based	o	o	o	o	o

Avoid testing on material deemed trivial	o	o	o	o	o
Ensure wording and sentence structure is succinct	o	o	o	o	o
Avoid hinged questions – questions that rely on answer from previous question	o	o	o	o	o
Abstain from stereotyping race, gender, or other factors/Screen for offensive content or scenarios	o	o	o	o	o
Support questions, answers, and rationales with research and/or other reputable evidence	o	o	o	o	o

Avoid the use of absolute terms (e.g. always, never, all, only)	o	o	o	o	o
Avoid use of items from commercial item banks	o	o	o	o	o
Create a pool of high-quality test items (faculty generated)	o	o	o	o	o

Best practices: Item Stem

For the following items, please honestly rate how frequently you use the practice when creating item stems. If you are unfamiliar with the practice, please choose "not familiar with the practice".

	Not familiar with practice	Do not use	Sometimes	Half of the time	Most of the time
Write stem as a question or partial statement	o	o	o	o	o
Explicitly state the information you are seeking; the question clearly defines the problem	o	o	o	o	o
Ensure that the stem can stand alone, is clear and complete (e.g., May be answered by competent student without	o	o	o	o	o

provided choices)					
Write items that have only one correct answer (except in cases where 'select all' is specified)	o	o	o	o	o
Avoid extraneous material not needed to answer the question	o	o	o	o	o
Avoid providing cues to the correct answer within the stem	o	o	o	o	o

Emphasize key words (e.g., best, priority, first) in *italics*, <u>underlined</u>, or **bolded**	o	o	o	o	o
Avoid the use of negative words in the stem (e.g., except, not incorrect)	o	o	o	o	o

Best practices: Item Distractor

For the following items, please honestly rate how frequently you use the practice in creating item distractors. If you are unfamiliar with the practice, please choose "not familiar with the practice".

	Not familiar with practice	Do not use	Sometimes	Half of the time	Most of the time
Ensure that all options are grammatically consistent with the stem	o	o	o	o	o
Arrange options in a logical order (e.g., numerical, alphabetical)	o	o	o	o	o
Ensure that all options equal in length and amount of detail	o	o	o	o	o
Verify that all options are mutually exclusive	o	o	o	o	o

125

(e.g., number ranges do not overlap)					
Evaluate that all options are plausible; options should be homogenous without obvious outliers in content	o	o	o	o	o
Resist presenting false information via distractor items – use all correct answers, forcing students to pick the "best" answer	o	o	o	o	o
Distribute correct answers randomly and evenly	o	o	o	o	o

Avoid the use of vague terms (e.g., frequently, often, occasionally) in all the options	o	o	o	o	o
Avoid the use of "all of the above" or "none of the above"	o	o	o	o	o
Consider three option items instead of 4-5 option items if an additional distractor is not plausible.	o	o	o	o	o

Best practices: Pre-Exam Review

For the following items, please honestly rate how frequently you use the practice in your pre-test review. If you are unfamiliar with the practice, please choose "not familiar with the practice".

	Not familiar with practice	Do not use	Sometimes	Half of the time	Most of the time

Proofread to ensure that answers/clues to one item are not provided elsewhere within the exam	o	o	o	o	o
Proofread exam for understandability and conflicts between questions	o	o	o	o	o
Corroborate with peer to complete review of exam items	o	o	o	o	o
Screen for writing flaws	o	o	o	o	o
Ensure instructions are concise, clear and not open to further explanation	o	o	o	o	o

Best practices: Post-Exam Review

For the following items, please honestly rate how frequently you use the practice in your post-exam review. If you are unfamiliar with the practice, please choose "not familiar with the practice".

	Not familiar with practice	Do not use	Sometimes	Half of the time	Most of the time
Utilize psychometric analysis to assure that the test is valid and internally consistent	o	o	o	o	o
Track difficulty index	o	o	o	o	o
Track item discrimination index	o	o	o	o	o
Track distractor efficacy	o	o	o	o	o
Track exam reliability (e.g. KR20)	o	o	o	o	o

Protect the integrity of the exam	o	o	o	o	o

Best practices: Item Revision

For the following items, please honestly rate how frequently you use the practice in item revising. If you are unfamiliar with the practice, please choose "not familiar with the practice".

	Not familiar with practice	Do not use	Sometimes	Half of the time	Most of the time
Use statistical data to revise item stems prior to next administration of exam	o	o	o	o	o
Use statistical data to revise item distractors prior to next administration of exam	o	o	o	o	o

Section 2: Facilitators and Barriers to Exam Item Best Practice Implementation

The literature suggests the following factors as facilitators to implementation of exam item best practices. Please select the **top three** facilitators that you feel would be most beneficial for your practice.

☐   Available faculty development funds

☐   Testing policy which includes exam creation best practices

☐   Faculty development regarding exam item best practices

☐   Department item developer position to support faculty

☐   Faculty within the department with expertise in item creation, exam evaluation, and item revision

☐   Peer feedback mechanisms

☐   Exam item mentoring for new faculty

☐   Administrative support for exam creation and revision *(e.g. administrative personnel who enter exam items into LMS or track item statistics)*

☐   Administration support for use of exam item best practices *(e.g. advocates for faculty knowledge development, time for exam item development, and funding to support this work)*

If you feel a facilitator of implementing exam item best practices was not included above, please include here:

_____

These factors were identified in the literature as barriers to implementation of exam item best practice. Please indicate if the factor acts as a barrier in **your** practice to the implementation of exam item best practices.

If you are unsure if the factor is a barrier to exam item best practice implementation, please select "No."

	Yes	No
Time constraints	o	o
Lack of faculty development regarding item creation	o	o
Lack of faculty development regarding exam post-test review	o	o
Lack of faculty development regarding item revision	o	o
Lack of peer support within department	o	o
Lack of administrative support *(e.g., administrative personnel who enter exam items into LMS or track item statistics)*	o	o

Lack of administration support *(e.g., does not advocate for faculty knowledge development, time for exam item development, and funding to support this work)*	o	o
Lack of adequate exam item tracking software *(e.g., ExamSoft, ExamView)*	o	o
Lack of testing policy that supports exam creation best practices	o	o
Lack of funding for faculty development	o	o
Lack of educational preparation for exam creation and evaluation	o	o

If you feel a barrier of implementing exam item best practices was not included above, please include here:

_____

Section 3: Faculty Demographics & Variables

What is your age?

- o 24 or younger
- o 25-34
- o 35-44
- o 45-54
- o Older than 55

What is your gender?

- o Male, including transgender men
- o Female, including transgender women
- o Prefer to self-describe as (non-binary, gender fluid, agender, etc. please specify)

  _____
- o Prefer not to say

What is the highest level of education you have obtained?

- o Associate's Degree
- o Bachelor's Degree
- o Master's Degree
- o Doctoral Degree

How many years since you completed your highest degree?

- o 0-2 years

- o 3-5 years

- o 6-10 years

- o Greater than 10 years

How many years of experience do you have teaching nursing?

o 0-2 years

o 3-5 years

o 6-10 years

o Greater than 10 years

At what type of institution do you work?

o Public university

o Private college or university

o For profit educational institution

o Other (please identify type of school0

_____

In which type of program do you teach? (Select all that apply)

☐   Associate's degree

☐   Bachelor's degree

☐   Master's degree

☐   PhD / DNP

How frequently do you write exam items?

o Weekly

o Monthly

o One to three times per semester

o Once to twice per year

o Never - standardized curriculum or program does not use tests

o Never - department has exam item writer

o Never - use commercial text bank questions without revision

---

Are you a Certified Nurse Educator (CNE)?

    o Yes

    o No

Does your nursing department have a testing policy that includes the best practices for exam creation? *If your testing policy contains some, but not all best practices, please choose "no"*

    o Yes

    o No

Does your department utilize exam blueprints as defined below?

Definition: document containing identified course and module objectives to be tested, the distribution of item type, cognitive level, and associated nursing process steps and/or client need categories If your blueprints contains some, but not all of the elements included in the definition, please choose "no".

    o Yes

    o No

Does your nursing department use software to track exam statistics (e.g. ExamSoft®, ExamView®)?

    o Yes

    o No

Do you receive administrative support for the creation of exams or tracking of exam results?

(e.g. administrative personnel who enter exam items into LMS or track item statistics)

   o Yes

   o No

Does your administration support faculty in the use of exam item best practices?

(e.g. advocates for faculty knowledge development, time for exam item development, and funding to support this work)

   o Yes

   o No

Does your institution provide funds for faculty development?

   o Yes

   o No

Do you have a nursing faculty mentor who helps you with exam item development?

(You may select both formal and informal mentor)

   ☐   Yes - formal mentor assigned by the department

   ☐   Yes - informal mentor

   ☐   No

Thank you for completing this survey.

To be entered into the drawing for one of the ten $75 Amazon.com gift cards, please follow the link to the google form. Your survey answers and the google form information will not be linked.

You may also indicate your desire to receive the executive summary at the conclusion of the research study on this form.

https://forms.gle/8y13qfiYSG4mbRWt5