

9-2018

Relationship Between Use of Collaborative Testing and Nursing Student Success

June A. Eastridge

Concordia University - Portland

Follow this and additional works at: <https://commons.cu-portland.edu/edudissertations>

Part of the [Education Commons](#)

CU Commons Citation

Eastridge, June A., "Relationship Between Use of Collaborative Testing and Nursing Student Success" (2018). *Ed.D. Dissertations*. 152.
<https://commons.cu-portland.edu/edudissertations/152>

This Open Access Dissertation is brought to you for free and open access by the Graduate Theses & Dissertations at CU Commons. It has been accepted for inclusion in Ed.D. Dissertations by an authorized administrator of CU Commons. For more information, please contact libraryadmin@cu-portland.edu.

Concordia University (Portland)

College of Education

Doctorate of Education Program

WE, THE UNDERSIGNED MEMBERS OF THE DISSERTATION COMMITTEE
CERTIFY THAT WE HAVE READ AND APPROVE THE DISSERTATION OF

June Annette Eastridge

CANDIDATE FOR THE DEGREE OF DOCTOR OF EDUCATION

Heather Miller, Ph.D., Faculty Chair Dissertation Committee

Aaron Deris, Ph.D., Content Specialist

Edward Kim, Ph.D, Content Reader

ACCEPTED BY

Joe Mannion, Ed.D.
Provost, Concordia University, Portland

Sheryl Reinisch, Ed.D.
Dean, College of Education, Concordia University, Portland

Marty A. Bullis, Ph.D.
Director of Doctoral Studies, Concordia University, Portland

Relationship Between Use of Collaborative Testing and Nursing Student Success

June Annette Eastridge
Concordia University–Portland
College of Education

Dissertation submitted to the Faculty of the College of Education
in partial fulfillment of the requirements for the degree of Doctor of Education in
Higher Education

Heather Miller, Ph.D., Faculty Chair Dissertation Committee

Aaron Deris, Ph.D., Content Specialist

Edward Kim, Ph.D., Content Reader

Concordia University–Portland

2018

Abstract

Collaborative testing allows students to work together on a test, discussing answer options and coming to consensus on the best response. Research has shown that it decreases test anxiety, increases learning and critical thinking skills, and allows students to practice collaboration and teamwork. However, schools of nursing continue to use traditional individual testing and high-stakes testing in order to prepare graduates to take the National Council Licensure Exam for Registered Nurses. This causal comparative study used existing gradebook data to explore the relationship between the use of collaborative testing and nursing student success. A positive relationship was found between collaborative testing and course success, without disproportionately increasing grades of lower performing students. Licensure exam pass rates for students who passed a course due to collaborative testing points were equivalent to those of students who passed the course regardless of collaborative testing. This study provides important information for nurse educators who would like to use testing as a learning tool.

Keywords: collaborative testing, group testing, cooperative testing, dyad testing, second testing, double testing, testing best practices, testing for learning.

Dedication

This work is dedicated to nurse educators everywhere, who always think first of the patients that will be cared for by the future nurses we are charged with developing.

Acknowledgements

I would like to take this opportunity to thank those in my life whose sacrifices have made my pursuit of higher education possible. First, my husband, Wes, who has encouraged all of my interests, even when it was not convenient. No one in either of our families has ever tried what I have sought to accomplish, and we certainly had no idea what we were getting into. Second, my children, Brandon, Deanna, and Erik, who have had years with fewer conversations with me than they would have liked. Last, my pampered puppy, who has spent less time in my lap.

Additionally, I would like to thank my dissertation chair, Dr. Heather Miller, who is an amazing and caring person that helped me navigate unfamiliar territory. She would have been a fabulous nurse. Dr. James Moore helped me move past a mental block that threatened to slow me down, and made completing this research possible. Last, I would like to thank Jeff Zuckerman, whose editorial skills and sense of humor brought pleasure to something that could have been tedious.

Table of Contents

Chapter 1: Introduction	1
Introduction to the Problem	1
Background, Context, History, and Conceptual Framework for the Problem.....	1
Statement of the Problem.....	4
Purpose of the Study	5
Research Questions.....	5
Rationale, Relevance, and Significance of the Study	6
Summary	8
Chapter 2: Literature Review.....	10
Introduction to the Literature Review.....	10
Conceptual Framework.....	17
Review of Research Literature and Methodological Literature.....	21
Synthesis of Research Findings	37
Critique of Previous Research	39
Chapter 3: Methodology	43
Introduction.....	43
Purpose of the Study	44
Research Questions, Hypothesis, and Operationalization of Variables.....	45
Research Design.....	47
Target Population and Participants	48
Instrumentation	49
Data Collection	50

Data Analysis Procedures	52
Limitations of the Research Design	54
Internal and External Validity.....	55
Expected Findings.....	55
Ethical Issues in the Study	56
Summary	57
Chapter 4: Results	59
Introduction.....	59
Description of the Sample.....	59
Summary of the Results	61
Detailed Analysis	61
Chapter Summary	66
Chapter 5: Discussion and Conclusion	68
Summary of the Results	69
Discussion of the Results	71
Discussion of the Results in Relation to the Literature.....	73
Limitations	76
Implication of the Results for Practice, Policy, and Theory	77
Recommendations for Further Research.....	78
Conclusion	79
References.....	81
Appendix A: IRB Approval	94
Appendix B: Statement of Original Work	95

List of Tables

Table 1. Gradebook Columns51

Table 2. Intercorrelation for Collaborative Testing (CT) Points Earned and Student Final Test
Average63

Table 3. Group Differences on CT Points, Individual Test Average, and Final Test
Average65

Table 4. NCLEX-RN First Attempt Pass Rates for Students with Course Pass/Fail Status Change
due to Collaborative Testing Points66

Chapter 1: Introduction

Introduction to the Problem

Schools of nursing rely heavily on traditional individual testing models and high stakes testing for assessment purposes (Bailey, Mossey, Moroso, Cloutier, & Love, 2012; Killingsworth, Kimble, & Sudia, 2015; National League for Nursing, 2012; Stonecypher & Willson, 2014). A survey of schools of nursing by the National League for Nursing in 2011 found that at least one third of nursing programs used high-stakes standardized testing scores to make student progression and graduation decisions (National League for Nursing, 2012). There is reason to believe that this number has continued to increase as schools of nursing attempt to keep student first-time National Council Licensure Exam – Registered Nurse (NCLEX-RN) pass rates high (Barton, Willson, Langford, & Schreiner, 2014; March & Robinson, 2015; National League for Nursing, 2012; Oermann & Gaberson, 2017). This emphasis on traditional use of testing for assessment of learning has prevented nursing faculty from using testing in ways that could best benefit students and extend learning (Harrison & Wass, 2016; Killingsworth et al., 2015).

Background, Context, History, and Conceptual Framework for the Problem

Bennett (1998), in a 20-year-old report for the United States Department of Education, accurately outlined the effect computers, access to the internet, and cognitive science would have on educational testing. At the time, testing was primarily administered in a traditional format: individually with paper and pencil for assessment of learning in a controlled, secure environment. Three evolutions were predicted for the future of testing. First, a change in testing format would use emerging technology infrastructures, and more closely adhere to principles of cognitive science in question construction and format. This evolution would better allow

standardized evaluation of student performance. Second, again based on improving use of technology, computer adaptive testing would individualize the testing experience, and better reflect the learning and critical thinking ability of each student. The use of audio and video artifacts embedded in tests would present a more accurate representation of real-life problems. In retrospect, Bennett's predictions have been amazingly accurate. A third-generation prediction would continue the evolution of testing. This generation would embrace collaboration during testing, real-world problem solving, and seamlessly incorporate assessment of learning into the curriculum. The incorporation of collaborative testing fits well within the predictions posited in this historical government report.

Years of data have demonstrated that an increased focus on testing for summative assessment has not led to improved learning (Spurlock, 2013; Warner, 2017). There is currently a movement to correct that trend, making education more student centered and restoring education's focus on learning (Harrison & Wass, 2016; Pugh & Regehr, 2016). High-stakes testing is re-envisioned as more than a measurement on which to base important decisions. Creative approaches like collaborative testing can lead to increased learning as students are expected to recall information and apply it in new ways, increasing student success and program completion (Bailey et al., 2012; Gordon & Pajagopalan, 2016; Oermann & Gaberson, 2017).

Based on constructivism, new knowledge is built upon prior learning and experiences. Knowledge building is a complex and dynamic process that heavily relies on context, and socio-linguistic interaction (Duane & Satre, 2014; Knowles, Elwood III, & Swanson, 2015). Constructivism is a central component of both adult and transformational learning theories (Knowles et al., 2015; Mezirow, 1991). Adults learn through active problem solving. This learning is increased when students are challenged through interaction with others in ways that

present alternative views. Self-reflection and analysis then occurs that assists the learner to reach a deeper understanding of learned concepts (Duane & Satre, 2014; Knowles et al., 2015; Mezirow, 1991).

The Gordon Commission formed in 2011 to create a model for modern education that prepares adults to enter a dynamic, technological, and collaborative workforce. They defined three strands of the educational process that should be applied by educators: curriculum, instruction, and assessment. Within this model, curriculum, instruction, and assessment are all tools to promote continued learning. Assessment is used as more than a measure of previous learning, but a vehicle to promote active continued learning (Gordon & Pajagopalan, 2016). Testing effect is a process of learning through knowledge retrieval that is experienced by the student during testing. Knowledge retrieval has demonstrated benefits above that of traditional studying for learning. Through testing, students learn to apply information in different ways which improves transfer of information across formats, contexts, and domains (Baghdady, Carnahan, Lam, & Woods, 2014; Lang, 2014; LoGiudice, Pachai, & Kim, 2015; Pugh & Regehr, 2016; Vogler & Robinson, 2016; Wiklund-Hörnqvist, Jonsson, & Nyberg, 2014).

Collaborative testing is one way that testing can be used both as a tool for continued learning and for assessment. Students are challenged to produce rationales to support their responses and spend additional time sorting through alternative answer options. Learners who do not recall information are presented with an additional opportunity to learn. Research has most often used the collaborative test as a second test following traditional individual testing. Documented benefits include increased learning, communication, and use of group process. Additionally, students using collaborative testing have reported decreased test anxiety (Duane & Satre, 2014; Gilley & Clarkston, 2014; Jang, Lasry, Miller, & Mazur, 2017; Larsen, Butler, &

Roediger III, 2013; LoGiudice et al., 2015; Parsons & Teel, 2013; Peck, Stehle Werner, & Raleigh, 2013; Rivaz, Momennasab, & Shokrollahi, 2015; Vogler & Robinson, 2016).

Although the idea of testing as a learning tool is not new, it has not been widely used in schools of nursing. The small number of research studies by nurse educators has explored the relationship between collaborative testing and student outcomes. A positive relationship was discovered between use of collaborative testing and both end-of-course final exams and standardized testing (Hanna, Roberts, & Hurley, 2016; Peck et al., 2013; Wiggs, 2011). Use of collaborative testing as a second test decreased the need for posttest reviews (Centrella-Nigro, 2012). Students rated collaborative testing favorably for learning, group process, and decreased test anxiety (Duane & Satre, 2014; Martin, Friesen, & De Pau, 2014; Peck et al., 2013). Nursing faculty in one program noted a relationship between collaborative testing and end-of-program success (Molsbee, 2013).

Statement of the Problem

The problem this study addressed is an emphasis on traditional use of testing for assessment of learning that has prevented nursing faculty from using testing in ways that could best benefit students and extend learning. Schools of nursing devote time and resources to assessment of student learning through traditional individual testing, including high-stakes testing used throughout nursing programs to guide progression and graduation decisions. However, research supports the use of testing for more than an assessment measure. Testing should be viewed as a formative process that provides the student with opportunities to critically think, as well as giving the student and educator information on what has been learned and what areas need development (Gordon & Pajagopalan, 2016; Oermann & Gaberson, 2017). Nurse educators require additional evidence that deviating from traditional use of individual testing will

not lead to decreased preparation of graduates for the NCLEX-RN. The following question emerged to guide the development of a plan for research that addresses the factors described in this chapter: Is there a positive relationship between the nontraditional use of testing, specifically collaborative testing, and nursing student success?

Purpose of the Study

The purpose of this study was to provide evidence that assists faculty to make decisions about the use of collaborative testing in schools of nursing. Testing currently remains primarily a measurement of learning in schools of nursing, even though assessment of learning can also be performed through dialog, activities, projects, and other assignments, (Bailey et al., 2012; Halstead, 2013; Kantar, 2014; Killingsworth et al., 2015; Wood & Ezebuihe, 2016). Testing could be re-envisioned as part of a formative process, or testing for learning, while still providing evaluative information to the educator and guiding future learning (Bennett, 1998; Gordon & Pajagopalan, 2016). This would require a change in the testing culture within schools of nursing based on a better understanding of how testing can successfully be used for continued learning. This study intended to discover the relationship between collaborative testing and student success.

Research Questions

Primary question. What is the relationship between the use of collaborative testing and nursing student success?

Subquestions.

1. What is the relationship between collaborative testing points earned and student success based on final course test averages in an advanced medical-surgical nursing course?

2. What is the difference in the total collaborative testing points earned between the lower performing students and higher performing students from a medical-surgical nursing course?
3. What is the difference in the overall School of Nursing NCLEX-RN first-time pass rates and the NCLEX-RN first-time pass rates of those who passed a course due to collaborative testing?

Rationale, Relevance, and Significance of the Study

Nursing research into testing has most often focused on test construction, security, and test item analysis (Donaldson & Gray, 2012; Killingsworth et al., 2015; Stillwell & Krautscheid, 2016; Stonecypher & Willson, 2014). The purpose of existing research into the use of nontraditional approaches to testing was to determine impact of testing on learning, communication, anxiety, and collaboration (Duane & Satre, 2014; Hanna et al., 2016; Martin et al., 2014; Parsons & Teel, 2013; Peck et al., 2013; Rivaz et al., 2015). This study provides additional information for nurse educators and nursing curriculum developers about nontraditional use of testing and how that use correlates with the success of nursing students.

Definition of Terms

Assessment for learning. Based on constructivism, tests are used to extend knowledge and problem-solving ability. Testing effect potentiates learning, and testing is integrated throughout a course. Tests provide both learning opportunities, and formal and informal evaluation of student performance (Gordon & Pajagopalan, 2016).

Assessment of learning. Tests are administered to students to measure learning of specific course content. The focus is measurement, and is limited to cognitive domains (Gordon & Pajagopalan, 2016).

Collaborative testing. Tests are administered to groups of students who work together to solve contextual problems using both past and present learning. Answer options are explored as rationales are discussed. Cognitive, affective, social, and situative learning domains are engaged (Oermann & Gaberson, 2017; Sandahl, 2009).

Constructivism. Previous learning and experience are used as a foundation for building new knowledge. Learning is contextual and dynamic, occurring continuously across human performance domains (Knowles et al., 2015).

Critical thinking. Solving problems and making judgements occur through assessment and analysis of multiple factors. Context is important for defining desired outcomes (Oermann & Gaberson, 2017).

Nontraditional testing. Testing is applied in creative ways in order to further students' knowledge and development. Measurement of achievement is displaced as the primary purpose (Bennett, 1998; Gordon & Pajagopalan, 2016; Sandahl, 2009).

Testing effect. A cognitive process in which information becomes a part of the student's knowledge structure through knowledge retrieval rather than traditional studying that uses knowledge encoding. Repeated knowledge retrieval makes learned information more accessible for future use and has both short- and long-term benefits (Foss & Pirozzolo, 2017; Pugh & Regehr, 2016; Racsmány, Szollosi, & Bencze, 2018; Yang & Shanks, 2018).

Traditional testing. Tests are administered to individuals in a secure environment to measure knowledge and learning (Gordon & Pajagopalan, 2016).

Assumptions, Delimitations, and Limitations

Assumptions within this study centered on the characteristics of the adult learner. The validity of the study depended on the assumption that adult learners are engaged, motivated, and

desire to learn. All learners, when given the opportunity, will continue to build upon existing knowledge. Adult self-motivation causes students to take advantage of every learning opportunity in order to improve critical thinking abilities. This goal orientation leads the learner to continually improve and transform, adopting the characteristics of the professional.

The focus of this study was limited to senior nursing students who may exhibit the characteristics of the adult learner to a greater degree than other students who are at different points in their education. Nursing students in their final semester will have had a considerable amount of clinical experience. This fact, added to having a greater opportunity to gain general life experiences, increased the chances that students were motivated and ready to apply critical thinking to real-life scenarios. This study was delimited to the use of an existing data set. The data was previously de-identified and did not present an opportunity to gain additional information.

Summary

Schools of nursing frequently use traditional individual testing for assessment of learning. Clinical situation based NCLEX-RN style questions are thought to prepare students to take the licensure exam after graduation. Testing is often conducted within a high-stakes context for student evaluation to limit progression of lower performing students who are more likely to negatively impact program NCLEX-RN first time pass rates (Barton et al., 2014; Kantar, 2014; Killingsworth et al., 2015; Oermann & Gaberson, 2017; Spurlock, 2013). This unique situation has prevented widespread adoption of creative uses of testing for learning, and created a need for nursing education focused research that analyzes the relationship between use of testing for learning and student success. This research was intended to contribute to the body of knowledge related to nontraditional use of collaborative testing for learning and give nurse educators

confidence to use testing in creative ways that further promote student learning and professional development, better preparing graduates for practice.

Chapter 2: Literature Review

Introduction to the Literature Review

In recent years, educators have increased their focus on assessment due to a demand for increased accountability for learning to governing bodies and to consumers of higher education. Organizations have been asked for evidence that students are receiving the quality education that is purchased, and that program graduation leads to employment in graduates' areas of study (Gordon & Pajagopalan, 2016; Oermann & Gaberson, 2017). The Gordon Commission, formed in 2011, worked for two years to examine the use of assessment in education (Gordon & Pajagopalan, 2016). Through interdisciplinary collaboration, the commission developed a blueprint for transforming assessment from solely a measure of achievement to a tool for continued learning and developing critical thinking skills. Recommendations from the Gordon Commission and the National League for Nursing (Halstead, 2013) included taking a multidimensional approach to assessing learning and using creative approaches for testing, such as collaborative testing, to support learning.

Organization of the review. This literature review provides a foundation for the need to transform the use of testing in nursing. The conceptual framework includes discussions of complex learning, collaboration, and critical problem solving through adult and transformational learning theories. In addition, collaborative testing is part of a dynamic pedagogy that combines curriculum, instruction, and assessment into a complementary and dynamic unit. I examine recent publications and research on testing in schools of nursing, on high-stakes testing, and on collaborative testing, which includes testing effect and problems of test anxiety both inside and outside of schools of nursing within higher education.

I began by searching for information on how testing is currently used in schools of nursing. Search terms included *testing*, *nursing education*, *testing practices*, *high-stakes testing*, and *test security*. A search for *testing recommendations*, and *testing best practices in nursing* found position statements on the National League for Nursing (NLN), the American Association of Colleges of Nursing (AACN) and the Institute of Medicine (IOM) websites. I explored testing for learning through nursing, education, and allied health literature. Key words for the primary topic of study included *collaborative testing*, *group testing*, *cooperative testing*, *dyad testing*, *second testing*, and *double testing*. All searches were done using the Cumulative Index for Nursing and Allied Health Literature (CINAHL), ProQuest, PubMed, and Education Resources Information Center (ERIC) databases. I also read books on testing in nursing, the future of testing in education, Knowles's adult learning theory, and transformational learning theory.

Collaborative testing. Collaborative approaches to learning are increasingly used in health care professions (Berndt et al., 2015; Bines & Jamieson, 2013; Mueller-Joseph & Nappo-Dattoma, 2013; Piper, 2016; Tanriverdi et al., 2017; Tolsgaard, Kulasegaram, & Ringsted, 2016). Collaborative learning is an active approach that is student-centered, and increases critical thinking. It benefits diverse groups of students, and better prepares them for a diverse workplace and life-long learning. Students believe that collaborative learning better prepares them to apply knowledge and skills within the clinical setting, helping to bridge the theory-practice gap (Bines & Jamieson, 2013; Piper, 2016; Tanriverdi et al., 2017). Students with a greater affinity towards group work benefit the most from collaborative learning approaches. Student attitudes and perceptions influence the efficacy of collaborative learning, and those who have a more positive perception towards collaborative learning demonstrate decreased anxiety, with an increase in personal accountability for learning (Gorvine & Smith, 2015; Mueller-Joseph & Nappo-Dattoma,

2013; Spiers et al., 2014). Although student perception and attitudes towards this learning approach vary, it has consistently led to improved learning for all students (Spiers et al., 2014). There are mixed results regarding whether collaborative learning benefits task oriented learning, like hands-on clinical skills. Although learning is initially increased through dialog and observation, there comes a point where individual hands-on practice is limited by group-based learning (Retnowati, Ayres, & Sweller, 2017; Tolsgaard et al., 2016). Simulation learning is increasingly used in nursing education in part due to the opportunity for students to collaborate and solve clinical problems. Unfolding case scenarios used in the classroom are a form of low-fidelity simulation. Active participation by students has been found to increase student confidence and learning (Berndt et al., 2015). Collaborative learning is most effective for higher level learning, including extension of knowledge, critical thinking, and problem-solving (Berndt et al., 2015; Bines & Jamieson, 2013; Kantar, 2014).

Nursing graduates step into roles within a complex health care system that require them to work side-by-side with other disciplines to promote positive patient outcomes. Developing communication and collaboration skills is an important outcome for many professional programs (AACN, 2014; IOM, 2010). Collaborative testing has been used to increase student collaboration and teamwork skills while enhancing nursing knowledge and critical thinking (Sandahl, 2009).

Collaborative testing allows groups of students to work together on a test, discussing answer options, and coming to a consensus on the best answer. Students may simply take a test in groups, but most often in collaborative testing students first take a test individually and then divide into groups to take the same test a second time together (Duane & Satre, 2014; Hickey, 2006; Parsons & Teel, 2013; Sandahl, 2009; Zipp, 2007). Benefits of collaborative testing include increased test scores and learning, decreased test anxiety, improved teamwork and

collaboration skills, increased motivation, and an increased ability to think critically (Sandahl, 2009).

Evidence supports the use of collaborative testing across disciplines (Dahlstrom, 2012; Duane & Satre, 2014; Gilley & Clarkston, 2014; Hanna et al., 2016; Hanshaw, 2012; Hanson & Carpenter, 2011; Martin et al., 2014; Meseke, Nafziger, & Meseke, 2010; Peck et al., 2013; Rivaz et al., 2015; Sandahl, 2009; Siegel, Roberts, Freyermuth, Witzig, & Izci, 2015; Srougi, Miller, Witherow, & Carson, 2013; Wiggs, 2011). However, research on collaborative testing in nursing education is limited (Sandahl, 2009), and only eight studies were found from the past five years from schools of nursing. Testing is most often used as a strict measure of learning as nurse educators attempt to prepare students to take the National Council Licensure Examination for Registered Nurses (NCLEX-RN; Barton et al., 2014; Halstead, 2013; Killingsworth et al., 2015; National League for Nursing, 2012; Oermann & Gaberson, 2017; Santo, Frander, & Hawkins, 2013; Spurlock, 2013; Stillwell & Krautscheid, 2016; Stonecypher & Willson, 2014). Recent studies of collaborative testing have focused on grade inflation related to higher collaborative testing scores (Centrella-Nigro, 2012), the impact of collaborative testing on long-term retention of material (LoGiudice et al., 2015; Martin et al., 2014; Rivaz et al., 2015; Vogler & Robinson, 2016; Zhang & Henderson, 2017), using testing as a study tool outside of the classroom (Wissman & Rawson, 2016), the effects of collaborative testing on anxiety (Bovee, 2016), and the impact of collaborative testing on program completion and NCLEX-RN success rates (Eastridge, 2014; Molsbee, 2013)

Context: Nursing education. Nursing education is in a state of transition. Nurses comprise the largest segment of the U.S. health care workforce, and demand for highly trained and competent nurses is increasing. Passage of the 2010 Affordable Care Act brought about

national initiatives that have challenged nursing educators to produce graduates capable of leadership roles in a rapidly changing health care system. A joint effort by the Robert Wood Johnson Foundation and the IOM led to a 2-year study of nursing and nursing education (IOM, 2010). The report provided an action plan that continues to direct the focus of nursing education today.

Currently, nurses enter practice at four levels: the licensed practical or vocational nurse (LPN/LVN), the associate degree prepared registered nurse (RN), the 3-year diploma RN, and the bachelor's degree prepared RN. Each level prepares nurses to assume different scopes of responsibility and competency. An increasingly complex health care system places more demands on the RN to function in leadership and management of patient care, system-wide quality improvement, health policy, evidence-based practice and research, teamwork, collaboration, and specialty-area specific competencies (AACN, 2014; Duane & Satre, 2014; IOM, 2010; Martin et al., 2014). With additional educations, nurses assume greater responsibility. Nurses are encouraged to become life-long learners, achieving higher levels of education, including master's and doctoral degrees to meet the challenges of nursing practice. The AACN (2014) reported that higher education for nurses leads to better patient outcomes. The AACN recommended transitioning to a minimum bachelor's degree level for entry into nursing practice in the future. Nurses should be prepared to practice to the full extent of their educational training and contribute a voice towards improvement of the U.S. health care system (IOM, 2010). The National Advisory Council on Nurse Education and Practice has determined that nurses increasingly need an ability to apply critical thinking and clinical problem-solving skills, work within interdisciplinary teams, contribute a broad knowledge of physical, social, and behavioral sciences, analyze and communicate data, and adapt to a changing environment. In

2013, only 55% of the nursing workforce was educated at a bachelor's degree level or higher. It has become the responsibility of every nursing educator to ensure that nursing program graduates are prepared to employ clinical reasoning in a collaborative environment (AACN, 2014).

Significance: Assessment of learning versus assessment for learning. Amid an increased focus on assessment of learning, a system of punitive accountability has developed. A greater dependence on the results of standardized testing for grading and progress has negatively affected program quality (Gordon & Pajagopalan, 2016; Oermann & Gaberson, 2017; Warner, 2017). However, judging institutions, programs, and educators based solely on standardized test scores overlooks student demographics and institutional mission (Sullivan, 2014; Warner, 2017). Standardized testing is not the best single indicator of educational quality because students in the United States come from many different ethnic, language, socioeconomic, and educational backgrounds that all effect test-taking abilities. Testing can better be used to extend learning (Abel & Roediger, 2017; Atabek Yigit, Balkan Kiyici, & Çetinkaya, 2014; Baghdady et al., 2014; Brewer & Unsworth, 2012; Cantor, Eslick, Marsh, Bjork, & Bjork, 2015; Carpenter, 2012; Chen & Chuang, 2012; Fernández Alemán, Carrillo de Gea, & Rodríguez Mondéjar, 2011; Foss & Pirozzolo, 2017; Freda & Lipp, 2016; Griswold, Overson, & Benassi, 2017; Iwamoto, Hargis, Taitano, & Vuong, 2017; Larsen et al., 2013; Pan, Pashler, Potter, & Rickard, 2015; Raupach et al., 2016; Schmidmaier et al., 2011; Tullis, Finley, & Benjamin, 2013; Wiklund-Hörnqvist et al., 2014; Yang & Shanks, 2018).

Significance: Use of testing in schools of nursing. Testing is a ubiquitous influence in nursing education. For example, many schools of nursing include admission exam scores as part of their admission criteria (Oermann & Gaberson, 2017). Multiple-choice tests are commonly used to test student learning within nursing courses (Bailey et al., 2012; Oermann & Gaberson,

2017). Additionally, standardized testing and high-stakes tests are used to ensure students are meeting benchmarks that guide progression and graduation decisions (Barton et al., 2014; Killingsworth et al., 2015; McClenny, 2016; NLN, 2012; Santo et al., 2013; Sullivan, 2014). Trials are currently underway that use simulation for high-stakes assessment (Rizzolo, 2015). Faculty use normed test scores to make decisions on student learning, progression, and readiness to take the NCLEX-RN, the results of which every school of nursing is evaluated for quality based on first-time pass rates of graduates. In turn, pass rates are often the criterion used to determine continued program approval or closure. Many programs have detailed testing policies to promote academic integrity and support high-stakes testing consistency in order to use these scores to base important decisions (Killingsworth et al., 2015; Stonecypher & Willson, 2014). Although assessment is also performed through clinical practice rotations, class projects, papers, activities, and discussions, most schools of nursing rely most heavily on classroom examinations and high-stakes standardized tests for summative evaluation (Bailey et al., 2012; Killingsworth et al., 2015; Oermann & Gaberson, 2017; Rizzolo, 2015; Wood & Ezebuihe, 2016).

The IOM (2010) recommendations for improving health care delivery involved strengthening connection of education to practice, teaching not only nursing knowledge but also how to use that knowledge based on real practice problems. Students should graduate from nursing programs ready to participate in a dynamic and collaborative health care environment, addressing patient needs from a holistic approach (Duane & Satre, 2014; Martin et al., 2014; Wood & Ezebuihe, 2016). A variety of assessment strategies are needed to ensure that nursing students meet clinical competencies and can safely apply nursing knowledge in clinical situations. This calls for a shift from primary summative assessment through testing to a broader assessment approach of clinical practice, critical thinking, and teamwork skills (Halstead, 2013;

Wood & Ezebuihe, 2016). A multidimensional approach to assessment, augmented using clear and detailed grading rubrics, switches the focus from purely evaluative testing and frees the educator to creatively use testing as an additional teaching/learning tool (Harrison & Wass, 2016; Pugh & Regehr, 2016). This approach is more student centered and eases the punitive, high-stakes environment that has long characterized testing in schools of nursing. Redefining testing as more than a summative evaluation tool brings the educator closer to practicing within dynamic pedagogy (Gordon & Pajagopalan, 2016).

Conceptual Framework

This study was based on a belief that adult learners build upon existing knowledge, and learning is a progressive and continual process. With this constructivist approach, adult learning occurs over time and in many contexts. Adults benefit from a variety of learning approaches, and assessment of learning cannot be contextually separated from the process of knowledge building (Gordon & Pajagopalan, 2016; Knowles et al., 2015). Viewing assessment as a continuation of learning is a holistic approach to education promoted by the Gordon Commission in its quest to find answers to problems encountered by educators as they prepare students to work in a dynamic, technological, and collaborative global environment (Gordon & Pajagopalan, 2016; Harrison & Wass, 2016; Pugh & Regehr, 2016; Wood & Ezebuihe, 2016). This study explored collaborative testing through the lens of constructivism. Constructivism describes a process of knowledge building that is learner centered (Duane & Satre, 2014; Knowles et al., 2015) and supports a shift towards dynamic pedagogy (Gordon & Pajagopalan, 2016) that is further supported by the characteristics of adult learners offered by adult and transformational learning theories (Merriam, 2001).

Constructivism. Context is an important element of constructivism. Learners build new knowledge based on prior learning and experience. Knowledge can be filtered by existing mental models, and information is selectively retained. To increase retention, the learning environment should closely reflect the environment in which the student will use the newly attained skill and knowledge (Knowles et al., 2015). A testing environment that uses collaboration to resolve clinical problems is reflective of the practice environment of nurses, and allows students to build and retain their knowledge. In addition, constructivism promotes using other students to challenge a learner's thinking and provide alternative views. This need for reflection and analysis ensures that the learner understands not only facts but the larger conceptual application of those facts to practice (Duane & Satre, 2014; Knowles et al., 2015).

Dynamic pedagogy. Gordon and Pajagopalan (2016) describe a “dynamic pedagogy” (p. 114) that combines three strands for learning: curriculum, instruction, and assessment. This model is based on research from cognitive and learning sciences about how people learn. The body of knowledge to be learned in any domain becomes the vehicle to achieve the true goal of modern education: to teach people to think. Learners must become autonomously motivated to find information and apply it to real-life problems, working collaboratively within a team. These are the abilities currently required of the educated health care professional (AACN, 2014; Duane & Satre, 2014; IOM, 2010; Martin et al., 2014; Wood & Ezebuihe, 2016).

One curriculum strand of dynamic pedagogy involves the body of knowledge to be learned within a discipline. Gordon and Pajagopalan (2016) suggest an approach for transmitting this information to the student through presentation of material and learning activities. Curriculum is designed to not only transfer knowledge but to provide students with opportunity to apply knowledge to context specific situations. Sound curriculum has built-in assessment

measures so students can self-evaluate and guide faculty decisions. (Gordon & Pajagopalan, 2016).

The instructional strand provides intellectual support to the student who is learning to think critically. It should address both cognitive and affective learning, helping the learner to develop “habits of mind” (Gordon & Pajagopalan, 2016, p. 129) that will lead to continued learning and sound reasoning. Educators are called on to model professional characteristics as learners develop their own professional practice.

The assessment strand is dynamic and evolving as new information is discovered. Assessment measures evaluate both how much knowledge the student has gained and how adept they are at using that knowledge. Assessment is multidimensional and provides ongoing feedback to both the student and educator. It also leads to deeper learning as students practice information retrieval and application to real-life scenarios (Gordon & Pajagopalan, 2016). Researchers have demonstrated the value of knowledge retrieval over knowledge encoding to enhance student learning. The power of knowledge retrieval increases the value of testing for learning within a dynamic pedagogy (Abel & Roediger, 2017; Carpenter, 2012; Freda & Lipp, 2016; Schmidmaier et al., 2011; Wiklund-Hörnqvist et al., 2014).

Adult learning. Historically, learning has been viewed in pedagogical terms. Only in the last century has adult education become a topic of theory, research, and discovery. Adult learning poses a complex set of problems related to individual background, motivation, setting, and personality (Merriam, 2001; Mezirow, 1991). In the 1920s the original question was whether adults were capable of new learning. From a behavioral and psychological perspective, researchers conducted basic memory tests under timed conditions. Younger learners scored

higher than older learners. However, in untimed conditions, the results indicated that adults up to age 70 were able to learn to the same extent as younger people (Merriam, 2001).

In the mid-1900s, intelligence (IQ) testing became popular. When IQ tests were applied to adults as well as children, the results were mixed. Adults scored better on some test components, and worse on others. However, score composites remained stable until late in life. Influencing factors were identified, such as socioeconomic background, level of education, additional training, and general health status. Unfortunately, most research that involved adults were conducted in laboratories under artificial conditions originally designed for testing children. Eventually, the questions being asked turned from whether adults could learn, because that had been demonstrated, to how adult learning differed from childhood learning (Merriam, 2001).

A prominent adult learning theorist, Knowles (see Knowles et al., 2015; Merriam, 2001), viewed the adult learner as self-directed, motivated, and growth-oriented. Critics have claimed that the same may be true for some children, and that not all adults have developed these traits because of social and cultural influences (Merriam, 2001). Although these arguments highlight the individuality represented by the diversity of adult learners, it does not exclude the inherent adult characteristics of maturity and goal pursuit that is found among nursing students. Knowles's model of adult learning provided a guide for nursing educators to connect learning with purpose, making it meaningful to reflect student goals. As directed by the AACN (2014) and IOM (2010), school of nursing graduates must be prepared to enter the profession ready to function in a complex health care system, working collaboratively to apply sound clinical reasoning (Duane & Satre, 2014; Martin et al., 2014; Wood & Ezebuihe, 2016).

Transformational learning. Mezirow (1991) extended what is known about adult learning by adding discussion of the importance of self-analysis and critical self-reflection,

dialog, making meaning, and consensus building. Congruent with constructivism, autonomous individuals increase their personal understanding through rational discourse when members of the group bring with them accurate and complete information, an openness to differing ideas, an ability to critically self-reflect, equal opportunity to present an argument, and the ability to weigh the evidence and assess all arguments objectively (Merriam, 2001; Mezirow, 1991). This takes the adult learner within a professional program from the realm of learning about a profession to learning to use a body of knowledge to function as a member of the profession. Learning activities can be designed to further this type of integrative and collaborative learning, including testing for learning (Duane & Satre, 2014).

Collaborative testing capitalizes on adult autonomy and self-direction, placing students in groups to discuss real-life problems, reflect on thinking and reasoning, make meaning, and collaboratively come to a consensus on appropriate clinical decisions. Adult learning theory and transformational learning provide direction for the educator who wants to use every opportunity to further professional development of the student.

Review of Research Literature and Methodological Literature

Use of testing in nursing education. Assessment in education involves collecting information about student learning to gauge the effectiveness of schools and educational programs. Assessment is used for evidence of the quality of educational services (Gordon & Pajagopalan, 2016; Oermann & Gaberson, 2017). Graduates from schools of nursing must pass the NCLEX-RN to practice as a nurse, and NCLEX-RN first-time pass rates are the primary measure used to assess program quality by state boards of nursing and national nursing education accreditation services. As such, nursing educators have emphasized the importance of frequent evaluation of student learning based on individual, NCLEX-style testing. (Halstead, 2013;

Kantar, 2014; Killingsworth et al., 2015; Oermann & Gaberson, 2017; Santo et al., 2013; Spurlock, 2013). The literature on testing in nursing has focused on NCLEX-style test item construction, test-item analysis, and test revision (Killingsworth et al., 2015) and test administration, test security, and academic integrity (Barton et al., 2014; Stillwell & Krautscheid, 2016; Stonecypher & Willson, 2014).

In addition to testing as a primary tool for assessment of learning (Kantar, 2014; Stillwell & Krautscheid, 2016; Stonecypher & Willson, 2014), educators have increased the use of high-stakes, NCLEX-RN style standardized exams, including exit exams, which require students to meet a benchmark score to graduate (Spurlock, 2013). The purpose of the NCLEX-RN is to prevent graduates who are not well prepared for practice from entering the nursing workforce. Increased use of testing to weed out students who are at risk of NCLEX-RN failure defeats the purpose of the licensure exam, artificially inflating pass rates for schools of nursing. While this approach prevents negative consequences from boards of nursing and accreditation agencies for low pass rates, it also can hide problems with curriculum, faculty, and instructional approaches (NLN, 2012; Spurlock, 2013). Program quality is better assessed using a combination of measures in addition to NCLEX-RN pass rates, including student characteristics, persistence, and graduation rates. Time from graduation to licensure exam should also be considered, because the longer students wait to test, the lower their chances of passing, a factor that nursing programs have no control over (Spurlock, 2013). These measures, combined with NCLEX-RN pass rates, provide a more complete picture of program quality and relieve the pressure on programs to ensure that all students test well (NLN, 2012; Santo et al., 2013; Spurlock, 2013)

In 2011, 20% of schools required a minimum set score on standardized exit exams (Halstead, 2013). By 2014, 44.4% of schools had a set benchmark for standardized testing that

students must meet to pass a course or graduate. Other schools (68.7%) used standardized testing as part of course grades. Most of these schools have a prescribed policy for remediation and retesting up to five times. From these schools, 85% included self-guided student preparation for standardized testing, and 53.7% used self-guided remediation prior to retesting. Fifty-six percent of schools had consequences for not meeting benchmarks involving failure of course, delay in graduation, and delay in sitting for the NCLEX-RN (Barton et al., 2014; Gibson, 2014). Because of the serious implications for students related to progression and graduation based on standardized test scores, schools must prepare students for testing, as well as provide clear resources for remediation. Students value standardized testing for learning; however, a high-stakes context presents a threat that diminishes learning potential (McClenny, 2016). Students must be well prepared for the potential outcomes of these tests and provided with avenues for increased learning and success (Barton et al., 2014; Halstead, 2013; NLN, 2012; Oermann & Gaberson, 2017; Santo et al., 2013).

The heavy reliance on test scores for assessing learning, combined with use of high-stakes standardized testing for progression and graduation decisions in schools of nursing, has led to a rise in test anxiety for nursing students. Test anxiety harms learning; conversely, addressing anxiety can lead to an increase in student test scores (Røykenes, Smith, & Larsen, 2014; Sullivan, 2014). Approaches to addressing test anxiety have included hypnotherapy, relaxation, visualization, music, and aromatherapy (Gibson, 2014). Student self-efficacy inversely affects levels of test anxiety, with higher self-efficacy leading to higher exam scores (Gibson, 2014; March & Robinson, 2015). However, even confident students can experience negative effects from test anxiety (Røykenes et al., 2014). Collaborative testing has been found

to decrease test anxiety for many students (Centrella-Nigro, 2012; Oermann & Gaberson, 2017; Parsons & Teel, 2013; Peck et al., 2013; Sandahl, 2009; Wiggs, 2011).

Cheating is a concern for nurse educators and influences testing practices. Because 21%-90% of college students admit to cheating (Stonecypher & Willson, 2014, p. 167), detailed behavioral and academic integrity policies are often in place to provide students with clear guidelines to deter cheating and create a culture of integrity. Stonecypher and Willson (2014) recommend that policies be visible and readily available to all students, and academic integrity policies be included in orientation for all incoming students. Test security practices should be consistently followed by faculty including scrambling of test questions and answer options on multiple test versions, active proctoring of exams, not allowing students to leave the testing area without an escort during an examination, refusal to answer student questions during an exam, and securing student's personal items away from the testing area (Stillwell & Krautscheid, 2016; Stonecypher & Willson, 2014). Types of questions faculty reported encountering most often during an exam involved clarification of unclear wording and word definitions. Faculty who did agree to answer questions did so to make students feel supported. "Examinations are used to evaluate individual student learning," (Stillwell & Krautscheid, 2016, p. 168), which adheres to the traditional use of testing solely for assessment.

Despite evidence to support a more student-centered approach to education, many curricula continue to follow the traditional practice of lecture followed by individual testing of content learning (Kantar, 2014). Kantar (2014) posited that these traditional practices do not encourage higher level learning or assist the learner to develop competencies needed in the 21st century workplace. One school of nursing was compelled to revise its traditional testing practices that were based on NCLEX-RN style exam questions when the school encountered a clinical

practice site with a policy against presenting health care workers with any type of misinformation, like that used in the distractors of multiple-choice testing (Bailey et al., 2012). Nursing educators researched the problem and found that psychology and education had long been concerned with a negative testing effect that enforces learner misconceptions and misinformation through this type of testing. Nursing faculty then developed a plan for assessment that incorporated a variety of evaluation strategies. Additionally, the faculty partnered with clinical practice sites to identify client and practice factors that should be included in nursing student evaluation. Thus, schools of nursing can successfully find alternatives outside of traditional testing models for student evaluation (Wood & Ezebuihe, 2016). Findings with this case study included an acknowledgement that, despite the evidence of negative testing effect, schools of nursing would not abandon their heavy reliance on multiple-choice, NCLEX-RN style testing. However, attention should be given to using well-constructed test questions based on sound psychometric principles (Bailey et al., 2012; Wood & Ezebuihe, 2016). Timely answer feedback and discussion of question and answer rationales are key to combating negative learning effects (Bailey et al., 2012; Peck et al., 2013).

Nursing programs are encouraged to expand their assessment practices to incorporate a broad range of approaches. For health care practice professions, it is especially important to weight evaluation of clinical practice appropriately to convey value of the practice component to the student (Donaldson & Gray, 2012). A need for multimethod approaches to assessment was found in nursing, midwifery, medicine, and allied health programs (Donaldson & Gray, 2012; Wood & Ezebuihe, 2016). Assessment of clinical practice can be affected by subjective factors such as student relationship with the evaluator and the design of the evaluation tool. It is important that evaluation of student practice performance be guided by clear and specific

guidelines to abate this effect, giving the nursing educator confidence to base student grades on measures outside of individual test scores (Donaldson & Gray, 2012).

Testing effect. Testing for learning has been addressed in health care programs. However, traditional use of testing as a summative assessment has created a culture that is difficult to change. Students view a test as a barrier to be overcome, and often are refractory to the learning opportunity that can occur unless the testing environment and methods of implementation are changed (Harrison & Wass, 2016; Pugh & Regehr, 2016). Feedback provided by instructors after a test is often used for competitive comparison or disregarded by those who passed an exam, as students look ahead to the next obstacle to be overcome (Harrison & Wass, 2016). Feedback is most effective when it comes from a nonthreatening source with whom the students feels a partnership. Linking testing to clinical practice situations while allowing students to co-create feedback and critically appraise their own thinking creates an environment that might alter the testing dynamic and encourage students to continue learning through the process of testing (Larsen et al., 2013; Pugh & Regehr, 2016).

Testing has been shown to improve learning, a phenomenon known as testing effect (Foss & Pirozzolo, 2017; Iwamoto et al., 2017; LoGiudice et al., 2015; Pugh & Regehr, 2016). Testing effect is a process where information better becomes a part of the student knowledge structure through retrieval over traditional studying. Knowledge retrieval makes learned information more accessible for future retrieval compared to knowledge encoding, and has shown to have both short- and long-term effects (Griswold et al., 2017; LoGiudice et al., 2015; Pugh & Regehr, 2016; Wiklund-Hörnqvist et al., 2014). In addition, students have demonstrated an ability to apply information in discrete ways when learning through knowledge retrieval. Students who learn through testing demonstrate improved knowledge transfer across test formats, contexts, and

domains. Improved learning is seen not only in information recall but also in situational application of knowledge, demonstrating students gain more than factual information through knowledge retrieval. They also gain an understanding of how that information can apply to real-life scenarios (Baghdady et al., 2014; Carpenter, 2012; Yang & Shanks, 2018). In a study involving fourth-year medical students, it was found that repeated testing enhanced clinical reasoning more than instructor led case studies (Raupach et al., 2016).

Exercising retrieval pathways has been found to decrease reaction times. Students who learn through testing become quicker with responses, and are able to use knowledge without exercising extreme attention. Problem-solving and applied skills are processed quicker, less effected by distraction, and retained long-term with repeated retrieval (Racsomány et al., 2018). Additionally, learning through retrieval is minimally affected by divided attention, whereas learning through studying, or memory encoding, is negatively impacted by distraction (Buchin & Mulligan, 2017).

Testing effect has primarily been studied through pretest–posttest design with different learning interventions applied. Testing as a learning intervention has consistently led to higher posttest scores (Abel & Roediger, 2017; Atabek Yigit et al., 2014; Freda & Lipp, 2016; Raupach et al., 2016; Schmidmaier et al., 2011; Tullis et al., 2013; Wiklund-Hörnqvist et al., 2014). An important factor in the use of testing for learning appears to be the availability of immediate feedback (LoGiudice et al., 2015; Wiklund-Hörnqvist et al., 2014). Despite concerns that testing for learning benefits low performing students more than others (Brewer & Unsworth, 2012), both lower and higher performing students can experience the benefits of increased learning through testing (Jang et al., 2017; Pan et al., 2015). In one study, marginal knowledge was reinforced through multiple-choice testing. This approach stabilized marginal knowledge as well as

restudying. Exposure to correct information with multiple-choice questions seems to stabilize not immediately recallable information, making it more retrievable for future use (Cantor et al., 2015). Testing effect is greatest when tests are separated by a few days or a week (Mulligan & Peterson, 2015; Wiklund-Hörnqvist et al., 2014). Delayed retesting also minimizes negative effects of testing produced by absence of immediate feedback and reinforcement of an incorrect answer selection (Mulligan & Peterson, 2015).

Testing for learning is being used more often in health care programs because of the enhanced capacity for deeper understanding and application of knowledge it provides (Baghdady et al., 2014; Freda & Lipp, 2016; Raupach et al., 2016). Schools of nursing have used test questions to promote learning outside of the classroom. One school of nursing used an online computer application to challenge students in a competition that involved answering multiple-choice questions. Students who participated in the additional online activity performed better on exams in the short-term. However, an exam given 10 weeks after the intervention showed no long-term gains (Fernández Alemán et al., 2011). Another school of nursing provided an intervention group with 20–30 question practice tests for each unit of study, and found that the experimental group scored significantly higher on a midterm exam. Final exam scores were not significantly different (Chen & Chuang, 2012). These short-term gains demonstrate that use of test questions to promote learning for the individual are immediately effective. However, the addition of collaborative discussion and feedback was not explored in these studies.

Test enhanced learning with self-generated explanations were shown to provide the most long-term benefits in retention and application of material when summative testing was provided six months after the application of four interventions to randomly assigned groups of students: testing with self-generated explanations, testing without explanations, studying a review sheet

with self-generated explanations, and studying a review sheet without explanations. Students who learned using testing with self-generated explanations scored significantly higher than students from other groups (Larsen et al., 2013). One hundred twenty-two sociology students demonstrated that collaborative testing as a second test improved long-term performance on a final exam given 10 weeks later. This approach, combining learning and assessment, was found to be successful for both students who had prior knowledge of the tested material and those who did not, and provided learning enhancement for all students (Zipp, 2007).

Collaborative testing. The benefits of testing effect have also been associated with collaborative testing. LoGiudice, Pachai, and Kim (2015), through a systematic review of the literature, report on both cognitive benefits and concerns when using collaborative testing. Through laboratory study, a phenomenon called “collaborative inhibition” (p. 379) occurs when individuals within a group are unable to recall to their full potential due to the distraction provided by discussion from other group members. “Production blocking” (p. 379) may also occur as group members wait their turn to speak. Providing individuals with time to answer questions alone prior to the group process offsets collaborative inhibition and production blocking (LoGiudice et al., 2015). Additionally, through “socially shared retrieval-induced forgetting” (p. 380), group members forget related but undiscussed information. When groups produce incorrect answers, negative learning can occur.

However, cognitive benefits to group testing also exist. Group members who do not recall information are exposed to and provided with another opportunity to learn. Cues allow groups to build upon what each member recalls, and produce correct answers. Counter to negative learning, students who incorrectly remember information can be convinced of the

correct information by members of their group. Retention of complex material is enhanced more through testing effect than simple knowledge (LoGiudice et al., 2015).

Collaborative testing is most often used as a second test after students answer questions individually (Centrella-Nigro, 2012; Dahlstrom, 2012; Duane & Satre, 2014; Eastridge, 2014; Gilley & Clarkston, 2014; Hanson & Carpenter, 2011; Heglund & Wink, 2011; Hickey, 2006; Jang et al., 2017; Leight, Saunders, Calkins, & Withers, 2012; LoGiudice et al., 2015; Molsbee, 2013; Parsons & Teel, 2013; Rivaz et al., 2015). Group discussion of the test, with opportunity to explore answer options and rationales, decreases the need for a posttest review. Arguing about answers and points for alternate answers is diminished with use of collaborative testing, making posttest reviews more pleasant for both students and educators (Centrella-Nigro, 2012; Eastridge, 2014). Students report through course evaluations that using collaborative testing for posttest review enhances learning of course content (Centrella-Nigro, 2012).

Groups can self-select or be randomly assigned. Self-selection is a common approach (LoGiudice et al., 2015), but, even in studies that used self-selection, authors have recommended random assignment to prevent students who know their group composition from dividing material among themselves for studying (Hickey, 2006; Siegel et al., 2015). This tactic is referred to as cognitive loafing, and is less likely when collaborative testing is administered as a second test (LoGiudice et al., 2015; Siegel et al., 2015; Srougi et al., 2013). Nafziger, Meseke, and Meseke (2011) studied two cohorts in a chiropractic course that used collaborative testing. One cohort tested with self-selected groups, and the other cohort tested in randomly assigned groups. Although no statistical difference was seen on overall exam scores, the randomized groups performed better on course quizzes. The authors noted that random assignment motivated all students to become proficient in all topics, while self-selection of groups possibly led to topic

focused studying. Evaluation of collaborative testing using focus groups with students who had experienced two semesters of testing in groups as a second test revealed that students preferred to test with the same group throughout the semester. This allowed for improved group dynamic, as students learned how to best work with the individuals within the group (Centrella-Nigro, 2012). Another study used quantitative gradebook analysis, and found that students perform better in groups than they do individually, and randomized groups perform better than student selected groups (Mbalamula, 2018).

Collaborative testing has led to concerns about academic dishonesty and grade inflation. For this reason, most educators use a combination of collaborative testing and individual testing (Jang et al., 2017; LoGiudice et al., 2015). By minimizing the impact of group scores on overall exam averages and combining low-stakes collaborative testing with higher stakes exams, educators can minimize inadvertent impacts on student grades (Centrella-Nigro, 2012; Duane & Satre, 2014). Communication between students during an exam is traditionally viewed as cheating (Stonecypher & Willson, 2014). Individual accountability is threatened when students test together. Collaborative testing as a second test mitigates this concern. However, there is still the concern that lower performing students will experience increased benefits through collaborative testing, without equal benefit for higher performing students. A study of 67 pre-medical and pre-engineering students in a calculus course used collaborative testing as a second test. Groups of four to five students were assigned groups to achieve diversity based on gender, grades, academic level, and pretest scores. Findings indicated that both lower and higher achievers who answered test questions incorrectly on their individual exam were able to provide the correct answer with collaboration. This finding persisted even when no one in the group had the correct answer individually (Jang et al., 2017). Traditional individual testing may lead to

increased academic dishonesty, especially when a test is high-stakes. Students faced with the need to individually meet a benchmark may feel compelled to resort to cramming and cheating out of desperation and a perceived lack of options (Pugh & Regehr, 2016).

Test anxiety is another topic addressed through collaborative testing. Students report that testing in groups decreases overall anxiety, and students feel motivated to prepare at least as much as they would have for an individual exam in order to perform well before their peers (Bovee, 2016; Hanshaw, 2012; LoGiudice et al., 2015; Meseke et al., 2010; Pandey & Kapitanoff, 2011; Parsons & Teel, 2013; Peck et al., 2013; Sandahl, 2009; Siegel et al., 2015; Wiggs, 2011). However, in one study using voluntary student surveys, only 36% of students reported decreased test anxiety with collaborative testing, while 21% reported it was ineffective or minimally effective in reducing anxiety (Duane & Satre, 2014). In a survey of 88 nursing students, LoGiudice (2015) reported that collaborative testing increased test anxiety because students found themselves second-guessing their understanding of material based on other's comments, similar to the phenomena described by LoGiudice as collaborative inhibition and negative learning. Students who have a higher preference for group activities experience a greater decrease in anxiety with collaborative learning (Gorvine & Smith, 2015).

Collaboration and teamwork skills are improved through collaborative testing. Students report improved ability to express their ideas both verbally and non-verbally. Students gain an ability to respectfully disagree, and become more assertive when stating a case for an answer they believe to be correct. Students feel that they become a team as they work towards a shared goal (Duane & Satre, 2014; Hanna et al., 2016; Parsons & Teel, 2013; Siegel et al., 2015). Group accountability is fostered as members learn that they cannot overly rely on the work of others. Each team member assumes responsibility for success of the group. When each student becomes

accountable, trust is developed, which increases group cohesiveness (Hanshaw, 2012). Students remain engaged in course topics throughout collaborative testing, even though they often become distracted or fail to participate during other collaborative learning activities (Duane & Satre, 2014). Students felt that this opportunity to practice teamwork and communication better prepared them for their future workplace (Martin et al., 2014).

Constructivist learning theory is applied through collaborative testing, compared to individual testing that uses an instructionist approach. The benefits of collaborative testing from constructivism include social and linguistic learning with discussion of questions and answer options, and knowledge construction as different learners contribute what they know to the conversation. Peer-to-peer learning is low pressure, relieving the anxiety that students may experience when reasoning out test answers alone (Duane & Satre, 2014).

Research results from schools of nursing are consistent with findings from other disciplines and demonstrate the value of collaborative testing for learning. When collaborative testing was used for remediation prior to a standardized exit exam, scores were not significantly higher, but fewer students fell below the benchmark that would require retesting (Hanna et al., 2016). Educators experienced less contention in posttest review when collaborative testing was used as a second test (Centrella-Nigro, 2012; Eastridge, 2014). Collaborative testing was highly rated by students for learning, group process, and decreased anxiety in most studies (Duane & Satre, 2014; Eastridge, 2014; Martin et al., 2014; Peck et al., 2013; Wiggs, 2011). Immediate learning effects were consistently positive (Duane & Satre, 2014; Eastridge, 2014; Martin et al., 2014; Peck et al., 2013); however, measures of long-term retention were mixed. While some studies demonstrated improved performance on future exams by those who participated in

collaborative testing (Rivaz et al., 2015; Wiggs, 2011), others did not (Martin et al., 2014; Molsbee, 2013).

Methodological review. Collaborative testing is a learning intervention and has been approached quantitatively in nursing education using quasi-experimental design (Martin et al., 2014; Rivaz et al., 2015; Wiggs, 2011). Student exam scores or final course grades have been compared with and without the use of collaborative testing. Student learning has been assessed using a control group that experiences traditional individual testing compared with an intervention group that has the addition of collaborative testing, most frequently as a second test (Heglund & Wink, 2011; Leight et al., 2012; Meseke et al., 2010). Students from previous semesters who have not experienced collaborative testing have also been used as the control in order to compare student exam scores after the application of the collaborative testing intervention (Hanna et al., 2016; Peck et al., 2013). Another approach is to compare each student's individual test score with their group score (Jang et al., 2017; Siegel et al., 2015). In one quasi-experimental crossover design, students rotated taking course exams individually or in groups, and a final test was used to assess if learning on topics tested through collaborative testing were better retained (Gilley & Clarkston, 2014).

Surveys have also been applied descriptively to the study of collaborative testing. Questions explore effects of collaborative testing on anxiety, study habits, teamwork and group process, communication, and perceptions of learning (Centrella-Nigro, 2012; Duane & Satre, 2014; Hanshaw, 2012; Hickey, 2006; Martin et al., 2014; Pandey & Kapitanoff, 2011; Parsons & Teel, 2013; Peck et al., 2013; Siegel et al., 2015; Srougi et al., 2013; Wiggs, 2011; Wissman & Rawson, 2016). Both faculty and students have been asked to evaluate their experiences with collaborative testing through completion of end-of-course surveys.

Methodological Issues

Quasi-experimental research on collaborative testing in nursing education uses convenience sampling, and most studies are limited to students in one course or graduating cohort. Small, nonrandomized samples limit the generalizability of findings and other factors that could influence research results—student demographics, experience of the educator, and other resources available to enhance student success—are not accounted for. In addition, when collaborative testing is added as a second test for an intervention group, the additional exposure to the exam that students who only experience individual testing do not have could account for improved student performance (Rivaz et al., 2015)

Much of the current literature on collaborative testing in nursing education is descriptive, addressing educator and student perception of the impact of collaborative testing on learning, communication, anxiety, and group process development (Centrella-Nigro, 2012; Duane & Satre, 2014; Hanna et al., 2016; Martin et al., 2014; Parsons & Teel, 2013; Peck et al., 2013; Rivaz et al., 2015; Wiggs, 2011). Students who have been exposed to collaborative testing successfully entering the nursing profession has been addressed only in one study (Molsbee, 2013). Topics of grade inflation and stronger students enhancing weaker students' success have been explored with the use of faculty and student surveys, and gradebook analysis (Centrella-Nigro, 2012; Molsbee, 2013; Parsons & Teel, 2013; Wiggs, 2011). The assumption is that passing a course due to collaborative testing points, or weaker students benefiting from the knowledge of stronger students, is a negative outcome that borders on cheating (Jang et al., 2017; Stonecypher & Willson, 2014). However, peer learning with various paths to success is an effective educational approach based on constructivism and what is known about adult learning (Knowles et al., 2015; Mezirow, 1991).

Only one extended study was found that compares course success due to collaborative testing, with program completion and NCLEX-RN success. Nursing faculty tracked students who passed courses due to collaborative testing points to determine the impact on program completion and NCLEX-RN pass rates (Molsbee, 2013). This study was conducted over a period of 5 year, and involved three cohorts with a total of 127 students. Findings indicated that students who passed one course based on collaborative testing points successfully completed the program and passed the NCLEX-RN. However, students who passed more than one course based on collaborative testing points experienced decreased chances of successful program completion and passing the NCLEX-RN. Statistical analysis was not used to determine the significance and generalizability of these findings.

Testing effect is a topic closely associated with collaborative testing as a learning activity. Providing further evidence of improved learning with the addition of collaborative testing, or testing effect, is the objective of most quasi-experimental studies on collaborative testing (Hanna et al., 2016; Martin et al., 2014; Peck et al., 2013; Rivaz et al., 2015; Wiggs, 2011). Aside from testing effect, group formation (Nafziger et al., 2011), and generational learning styles (Hanson & Carpenter, 2011) were other topics of focus in nursing studies on collaborative learning. Other variables that impact student learning received little attention.

Ethical issues can influence choice of research approach. Stratifying students based on race or ethnicity, socioeconomic status, or gender related to benefits of collaborative testing would be more complex (Creswell, 2014). One study addressed differences in learning through collaborative testing based on gender, and found that both men and women benefited equally (Jang et al., 2017). Additionally, because testing effect and collaborative testing have been

demonstrated beneficial to students, withholding this intervention from a control group could result in student complaints.

Synthesis of Research Findings

Schools of nursing follow a traditional model of lecture or learning activity, followed by individual testing to evaluate learning (Duane & Satre, 2014; Kantar, 2014; Oermann & Gaberson, 2017). A driving concern for nursing educators is maintaining high first-time NCLEX-RN pass rates in order to satisfy state board of nursing regulations and accreditation agency expectations (Halstead, 2013; NLN, 2012; Oermann & Gaberson, 2017). This pressure has caused nurse educators to emphasize maintaining a secure and structured environment for individual testing (Barton et al., 2014; Stillwell & Krautscheid, 2016; Stonecypher & Willson, 2014), as well as an increased use of high-stakes standardized testing to guide student progression and graduation decisions (National League for Nursing, 2012; Santo et al., 2013; Spurlock, 2013). This approach has long served the needs and purposes of nursing education units.

A more student-centered approach places individual student development as the central concern. A change from an instructionist approach to a constructivist approach leads to use of testing as a continuation of learning. Collaborative testing provides opportunity for the student to increase knowledge and critical thinking through testing effect, and practice collaboration and teamwork skills while working together to solve patient-centered problems (Duane & Satre, 2014). This approach to learning within a professional program is also supported by adult and transformational learning theories that posit students are self-motivated and goal oriented, building understanding through dynamic interaction, critical discussion, and self-reflection (Knowles et al., 2015; Mezirow, 1991).

A focus on assessment of learning has threatened the use of assessment for learning (Harrison & Wass, 2016). Cramming and cheating are an outcome of traditional testing models and prevents real learning (Pugh & Regehr, 2016). Collaborative testing supports development of critical thinking and higher-level learning. Lower performing students may benefit the most from collaborative testing, and students' success and retention are increased as students go forward to perform better individually after experiencing group testing (Dahlstrom, 2012).

Standardized high-stakes testing will continue to play a primary role in determination of NCLEX-RN readiness because of the widespread adoption of policies that use standardized testing benchmarks to govern progression and graduation decisions (Barton et al., 2014; Halstead, 2013; Santo et al., 2013). These tests are associated with elevated levels of student stress and anxiety. Students should be prepared for testing based on questions that require application of knowledge to patient-based scenarios. Practice of critical thinking within groups that requires collaboration among diversely thinking students leads to improved critical thinking for the individual and better preparation for high-stakes standardized exams (Barton et al., 2014; LoGiudice et al., 2015).

Collaborative testing provides many benefits, including increased learning through testing effect, decreased test anxiety, application of teamwork and collaboration skills, use of group process, critical self-reflection, analysis of thinking, and, because of immediate feedback, decreased negative testing-effect (Baghdady et al., 2014; Hanna et al., 2016; Peck et al., 2013; Raupach et al., 2016; Sandahl, 2009; Siegel et al., 2015). Faculty concerns include the possibility of students cheating (Jang et al., 2017), and lower performing students passing a course based on the abilities of higher performing students (Parsons & Teel, 2013). A larger concern revolves around the impact a change to an assessment for learning culture (Gordon & Pajagopalan, 2016;

Harrison & Wass, 2016) would have on program graduate first-time NCLEX-RN pass rates (Eastridge, 2014; Molsbee, 2013).

Critique of Previous Research

Previous research has primarily focused on uncovering the benefits of collaborative testing related to student learning, communication and collaboration skills, and test anxiety (Sandahl, 2009). Student learning has been measured through comparison of individual and group exam scores, demonstrating that collaborative testing positively impacts learning, and is well liked by both students and educators (Centrella-Nigro, 2012; Duane & Satre, 2014; Hanna et al., 2016; Parsons & Teel, 2013; Rivaz et al., 2015; Warner, 2017). This focus has remained course centered and ignores the larger consequence of student graduation and NCLEX-RN pass rates.

Research on collaborative testing has involved variables that make systematic comparison difficult. Group sizes, self-selection or random assignment of groups, use in conjunction with individual testing as a pretest or posttest, or use as a stand-alone assessment, vary throughout previous research (LoGiudice et al., 2015). Collaborative groups generally consist of three to six students (Bovee, 2016; Duane & Satre, 2014; Gilley & Clarkston, 2014; Hanna et al., 2016; Jang et al., 2017; LoGiudice et al., 2015; Martin et al., 2014; Rivaz et al., 2015; Srougi et al., 2013). Although collaborative testing is most often used as a second test (LoGiudice et al., 2015), it can also be used alone (Hanshaw, 2012).

Only one study took the macro view of collaborative testing's impact on nursing students' successful entry into the profession (Molsbee, 2013). This study tracked students who passed courses based on collaborative testing scores to see how they did on end-of-program standardized testing and passing the NCLEX-RN. Students who passed one course based on

collaborative test scores could successfully complete the program and pass the licensing exam (Molsbee, 2013). Students who passed more than one course due to collaborative testing often failed or dropped out of the program. This failure to complete may indicate that lower performing students can get through courses without adequate learning, leaving them unable to continue as subsequent courses build upon knowledge attained in earlier courses. It may also uncover students who are struggling with factors outside of school that influence their ability to succeed. The factors that contribute to the observed findings are not clear. In addition, statistical analysis to determine significance of the finding was not performed.

Consequently, researchers have not measured the impact of collaborative testing on successful student completion and entry into the profession. This unanswered question has the likelihood of leaving nurse educators hesitant to use collaborative testing extensively.

Summary

An increasingly complex health care environment has led to recommendations from both the IOM and the AACN to transform nursing education. A more practice-centered approach is needed to prepare the contemporary nurse to assume added responsibility in management and coordination of patient care. Nursing educators are challenged to use educational approaches that better prepare the nursing graduate to step into a professional role. This has led to an increased emphasis on communication and collaboration skills within health care programs. Through collaborative testing nursing students can discuss practice-based questions and make decisions about patient care and professional nursing practice. Meaningful dialog and reflection on response options with analysis of rationales leads to higher level learning.

A review of the literature has shown that schools of nursing rely heavily on traditional and high-stakes testing for assessment of student learning and to make progression and

graduation decisions. However, a constructivist approach to adult and transformational learning supports the use of testing for more than assessment of learning. Testing is effective for continuation of learning and development of critical-thinking skills in application of knowledge. Research on collaborative testing supports the added benefits of decreased test anxiety, practice of communication and collaboration skills, use of group process, self-reflection, analysis of thinking, and diminished negative testing effects.

Collaborative testing has demonstrated value in promoting increased learning, with positive responses from both students and faculty. Increased learning through testing effect has been verified multiple times. However, there is a paucity of information on the relationship between use of collaborative testing and program completion and NCLEX-RN pass rates. Nursing program quality continues to be judged based on first-time NCLEX-RN pass rates, so nursing educators should know that students who pass courses as a result of collaborative testing have good program completion and NCLEX-RN pass rates. Without these data, it is unreasonable to expect nursing educators to deviate from use of traditional testing as the primary tool for assessment of student learning. Data are needed to support a change from an assessment of learning culture to an assessment for learning culture.

This review of the literature and a conceptual framework focused on a constructivist approach aims to better understand the benefits of collaborative testing. The review suggests examining the relationship between collaborative testing and student success including NCLEX-RN pass rates would provide socially significant findings. I can, therefore, claim that the literature review has provided strong support for pursuing a research project to answer the following research questions: What is the relationship between collaborative testing points earned and student success based on final course test averages in an advanced medical-surgical

nursing course? What is the difference in the total collaborative testing points earned between the lower performing students and higher performing students from a medical-surgical nursing course? Finally, what is the difference in the overall School of Nursing NCLEX-RN first-time pass rates and the NCLEX-RN first-time pass rates of those who pass a course due to collaborative testing?

Chapter 3: Methodology

Introduction

The use of testing for learning is based on a constructivist belief that adult learners continuously and progressively build upon existing knowledge. Adult learners enter higher education at different levels of ability based on their background, motivation, and personality. This is especially true in diverse classrooms with students of different ages, genders, educational preparation, and cultures. Promoting the success of diverse students requires creative approaches to learning and assessment, combining both to allow students to meet course outcomes (Knowles et al., 2015; Merriam, 2001; Mezirow, 1991). Collaborative testing is one creative approach to combining learning and assessment for adult learners that has shown promise for increased student learning. Through collaborative testing, students can work together on a test, discussing questions and answers in order to determine the best response (Sandahl, 2009; Zipp, 2007).

Collaborative testing has been used in schools of nursing to increase learning and critical-thinking skills, increase communication and collaboration, and decrease test anxiety (Duane & Satre, 2014; Martin et al., 2014; Peck et al., 2013; Sandahl, 2009) There is a paucity of research on the relationship between collaborative testing and overall course success (Molsbee, 2013). Further, because of the lack of empirical evidence about use of collaborative testing related to program NCLEX-RN pass rates, this creative testing model has not been widely adopted.

This chapter provides the purpose of this study, including the research questions (RQ), hypotheses (H), research design, variables, sampling, data, and methods for data analysis. Limitations of the study, including issues of internal and external validity, is also discussed. Finally, expected findings are stated, along with any ethical issues impacting the study.

Purpose of the Study

The purpose of this causal comparative design research was to test hypotheses derived from the principles of adult and transformational learning theories by examining the relationship between collaborative testing and nursing student success (Adams & Lawrence, 2014; Creswell, 2014). A review of the literature has found one nursing text (Oermann & Gaberson, 2017) and five peer-reviewed journal articles (Halstead, 2013; Kantar, 2014; Killingsworth et al., 2015; Stillwell & Krautscheid, 2016; Stonecypher & Willson, 2014) suggesting that nursing faculty are tied to a traditional use of individual testing for assessment. This traditional approach aims to prevent students who have trouble passing NCLEX-RN style tests from negatively affecting program NCLEX-RN pass rates. Need for traditional testing to screen for NCLEX-RN readiness has limited the use of collaborative testing for learning, and prevented the NCLEX-RN from fulfilling its purpose of barring licensure for unprepared candidates (Barton et al., 2014; Halstead, 2013; National League for Nursing, 2012; Santo et al., 2013; Spurlock, 2013; Sullivan, 2014).

This causal comparative research was designed to analyze existing gradebook data that provided values for total collaborative testing points earned, final course test averages, and limited NCLEX-RN success information based on seven semesters of one advanced medical-surgical nursing course, looking for relationships between these existing data. This course was taught in a School of Nursing on a rural university campus. Data represent student scores in the advanced medical-surgical nursing course from Fall semester, 2008, to Fall semester, 2011. All data were extracted without student names or other identifying information, and were determined to require no review by the Institutional Review Board (see Appendix A for supporting communication). Scores representing students who passed the course based on the addition of

collaborative testing were identified, with consideration of their final impact on program NCLEX-RN pass rates through appraisal of their end-of-program NCLEX-RN results. This research has the potential to either reinforce current traditional testing practices, or alleviate nurse educator fears related to use of testing as a learning tool.

Research Questions, Hypothesis, and Operationalization of Variables

RQ1. What is the relationship between collaborative testing points earned and student success based on final course test averages in an advanced medical-surgical nursing course?

H₀. There is no relationship between earned collaborative testing points and final course test averages in a medical-surgical nursing course.

H_a. There is a relationship between earned collaborative testing points and final course test averages in a medical-surgical nursing course.

Operationalized variables for RQ1. Two discrete variables were used. The predictor variable (x) was the total collaborative testing points each student earned during the course. The criterion variable (y) was the end-of-course test average each student achieved based on a student's individual test scores combined with collaborative testing points awarded according to the overall group score. These paired bivariate variables were analyzed for a correlative relationship (Adams & Lawrence, 2014).

RQ2. What is the difference in the total collaborative testing points earned between the lower performing students and higher performing students from a medical-surgical nursing course?

H₀. There is no difference in the average collaborative testing points earned between the lower and higher performing students.

H_a. There is a difference in the average collaborative testing points earned between the lower and higher performing students.

Operationalized variables for RQ2. The data were categorized into two independent groups based on end-of-course test averages. Those below the course pass benchmark of .75 represented the lower performing students and above .75 end-of-course test averages represented the higher performing students. These groups signified the independent variables. The dependent variables were the total collaborative testing points earned.

RQ3. What is the difference in the overall School of Nursing NCLEX-RN first-time pass rates and the NCLEX-RN first-time pass rates of those who pass a course due to collaborative testing?

H₀. NCLEX-RN pass rates for students who pass an advanced medical-surgical nursing course due to collaborative testing points are not significantly different than those of their graduating cohorts.

H_a. NCLEX-RN pass rates for students who pass an advanced medical-surgical nursing course due to collaborative testing points is significantly different than those of their graduating cohorts.

Operationalized variables for RQ3. The NCLEX-RN pass rate for students who passed a medical-surgical nursing course due to collaborative testing points is compared with the NCLEX-RN pass rates of the data set, based on the data for which these results are recorded. Two categorical variables were used: pass or fail of the course when collaborative testing points were not added to individual test scores, and pass or fail of the NCLEX-RN on first attempt after program completion.

Research Design

Causal comparative design for proposed study. Selection of research method was based on research questions that were identified through a review of the literature. Relationships between independent and dependent variables are explored through causal comparative research, including examining differences between groups. In this causal comparative (ex post facto) research, I used data from existing gradebooks to identify relationships between variables. Causal comparative research looks for relationships between the independent and dependent variables, and is ideal for use with existing data sets. Comparison between groups is also made using this research method (Field, 2005). Correlation statistic was employed to discover patterns within data. Researchers cannot control or manipulate variables using existing data, so even though relationships may be discovered based on a cause that has already been applied, absolute causation cannot be ensured (Creswell, 2014). When an educator observes a phenomenon, a causal comparative study can help confirm or repudiate that observation (Adams & Lawrence, 2014).

The choice of this causal comparative research design was prompted by the observation that students could pass an advanced medical-surgical course with the addition of collaborative testing points where they would have failed if only traditional individual testing were used. Originally, this observation led to informally tracking students beyond the course to see if they would then not only successfully complete the nursing program but also pass the NCLEX-RN on first attempt. Most, but not all, were able to complete and obtain their initial RN license. This study is designed to extend original informal instructor observations to a formal empirical analysis in order to quantify the observed relationship. The results contribute to a better understanding about the use of collaborative testing.

Support for causal comparative design from previous research. The choice of causal comparative design was supported by the literature. Non-experimental research designs have previously been employed by nurse educators to explore relationships between collaborative testing as an independent variable, and student outcomes as a dependent variable. Previous researchers found a positive relationship between use of collaborative testing and both end-of-course final exams and standardized testing (Hanna et al., 2016; Peck et al., 2013; Wiggs, 2011). A relationship between collaborative testing and end-of-program success was previously noted by nursing faculty in one program, with success in a single course due to collaborative testing correlating positively with program completion and student success on NCLEX-RN. However, success in more than one course due to collaborative testing correlated with failure to complete or NCLEX-RN failure (Molsbee, 2013). No descriptive statistical analyses were performed with the Molsbee (2013) study. In this study, I sought to extend knowledge of any relationship between collaborative testing and nursing student success.

Target Population and Sampling Method

The target population for this study is nursing students through use of existing school of nursing course data. As described earlier, these data were preexisting individual test scores, individual test scores with the addition of collaborative testing points, and NCLEX-RN results. No individuals were named because of the de-identified state of the existing gradebook data. Data represents 131 nursing students who were enrolled in an advanced medical-surgical nursing course that incorporated collaborative testing over a period of seven consecutive semesters.

The sampling method was a census of these data used from existing gradebooks that contained all test scores, both individually and with added group points, for the full seven semesters. Out of the 131 student records, only 40 included NCLEX-RN results. All 40 were

used in the final analysis, comparing NCLEX-RN pass rates for those who passed the course due to addition of collaborative testing point, and those who would have passed the course without the addition of collaborative testing.

Institutional Review Board (IRB) approval was first sought to use existing data for this causal comparative research in May 2017 (see Appendix A). The original academic institution response indicated that no IRB review was required for de-identified gradebook data. A full review application was submitted to Concordia University-Portland IRB in February 2018. Notification of approval through expedited review was received in March 2018.

Instrumentation

Preexisting, de-identified gradebook data represent nursing student scores on tests in an advanced medical-surgical nursing course. Each test contained 50 to 100 multiple choice or multiple-select NCLEX-RN style questions. Tests were first taken individually, and the score were recorded in a first gradebook column. The same tests were again taken in groups. This second test was scored, and points were added to each individual score based on the percentage correct earned by the group. A second gradebook column contained this second score. This second gradebook column, including collaborative testing points, was used to calculate the student's final test average for the course and their final course grade.

NCLEX-RN style questions are either clinically based, including both a nurse and a client, or require some form of clinical-reasoning (National Council of State Boards of Nursing, 2015). Students had a time limit of one minute per question for each individual measure. Collaborative testing that followed each individual test was not timed, but it typically required one third to one half of the time needed for the individual component. For example, a 50-question test was completed by the group in 15 to 25 minutes. Updating or improving measures

to maintain currency and clarity was done each semester, but otherwise measures remained consistent throughout the data collection period.

Data Collection

Preexisting data representing seven semesters of an advanced medical-surgical nursing course were used. Students in this course completed six individual tests over the semester, except for one semester that involved five tests. After each individual test, each student was randomly assigned to one of (*N*) groups by drawing a group number from a bowl. Each group contained four to six students, and class size ranged from 13 to 25 students. Thus, each student likely collaborated with a mix of higher and lower scoring test-taking students. After each student had completed the test individually, the students would then take the same test a second time with their groups. Collaborative testing was conducted after each of the individual tests and the group formation process was reinforced by peer-reviewed literature to support fair testing practices and decrease group homogeneity (Hickey, 2006; Siegel et al., 2015).

Each student had two scores for each of the course tests: an individual score and a score with the addition of collaborative testing. The collaborative testing score was determined by adding percentage points to the student's individual score based on group performance taking the same test a second time together. The final course grade was determined using the test scores that included the collaborative testing component. Individual scores were tracked only for later comparison. Each test consisted of 50 to 100 multiple-choice or multiple-select NCLEX-RN style questions.

The 12 scores were recorded in a gradebook for each semester as shown in Table 1. The individual score was adjusted based on group performance to attain the collaborative testing score. For groups that scored 90%–99% correct, 2% was added to the individual score to obtain

the collaborative testing score used to calculate the student’s final test average and course grade. For groups that scored 100% correct, the individual score was increased by 5% to determine the collaborative testing score used to calculate the student’s final test average and course grade.

Table 1

Gradebook Columns Providing Individual Score (I_Score) and With Collaborative Testing Points Added (C_Score).

	Test 1		Test 2		Test 3		Test 4		Test 5		Test 6		Final Average	
	I_Score	C_Score	I_Score	C_Score										
1	98	103	96	101	92	97	100	105	92	94	93.18	95.18	95.20	99.20
2	70	70	76	81	80	85	76	81	62	62	81.82	83.82	74.30	77.14

Details for the collaborative testing process were established through collaboration with other nurse educators during and after July 2008, beginning at a professional nurse educator conference. It was determined that the student had to have an individual score of 75% or greater to qualify for collaboration points, and that the 2% and 5% adjustments would provide ample incentive for participation without greatly inflating test scores.

The data set for this study comprised the de-identified final gradebook for each of the seven semesters. Each gradebook had the individual score and the collaborative testing score for each course test, and two columns containing the mean of the individual score and mean of the collaborative testing score. In addition, the total number of group percentage points earned over the semester was recorded for each student. Student success or failure on the NCLEX-RN exam postgraduation was provided by quarterly state Board of Nursing NCLEX-RN reports provided

to the nursing program. This information was placed in an additional gradebook column for those students who had passed the course due to collaborative testing points based on a required test score average of 75%, but who would have failed the course based on an individual score mean of less than 75%.

Prior to each collaborative testing experience, students were provided instructions. Within each group, students were to organize their approach. Some assigned a reader and a recorder, and some groups had members take turns reading questions. Active discussion was encouraged, but discussion between different groups was not permitted. Students were not allowed to have any outside resources, cellphones, or notes during the examination.

Data collection took place between August 2008 and December 2011 on a small campus of a major state university in the southwestern region of the United States. The campus served a greater than 90% minority student population comprising tribes and pueblos of American Indians, Hispanic/Latino, Asian, Middle Eastern, and other students for whom English is a second language (University of New Mexico-Gallup, 2017).

Data Analysis Procedures

Three separate data analysis procedures were used for this study. Pearson's correlation tested for a relationship between collaborative testing points and student success. Comparison of groups was done using both an independent sample *t* test, and chi-square.

RQ1. What is the relationship between collaborative testing points earned and student success based on final course test averages in an advanced medical-surgical nursing course?

Both collaborative testing points (independent variable) and final course test averages (dependent variable) are continuous variables. As such, I used Pearson's correlation for this RQ.

RQ2. What is the difference in the total collaborative testing points earned between the lower performing students and higher performing students from a medical-surgical nursing course? Correlative analysis of the lower performing half (Group A) and the higher performing half (Group B) of students were used to determine if there was a difference between groups for collaborative testing points earned. Because the two groups (independent variable) were categorical and earned collaborative testing points (dependent variable) were continuous, an independent sample *t* test was employed to test if there was a statistically significant difference between the two groups.

RQ3. What is the difference in the overall School of Nursing NCLEX-RN first-time pass rates and the NCLEX-RN first-time pass rates of those who pass a course due to collaborative testing? Chi-square test for independence was used to compare two groups based on nonparametric data. The NCLEX-RN pass rates of nursing students who passed a medical-surgical nursing course due to collaborative testing represented the first group, and overall pass rates of all students who graduated from the school of nursing during the time of data collection were the comparison group. Chi-square was used with two categorical variables to compare the observed frequencies found in the data, with the frequencies expected within the population. In this study, the test was used to look more closely at the NCLEX-RN passing rates of those who passed the course due to collaborative testing points.

The *p*-value of 0.05 for statistical significance was used for each analysis. Statistical significance provides no information about the importance of a correlation or difference regardless of whether it is statistically significant or not. Effect size is an indicator of importance and was reported in conjunction with the statistical significant results as recommended by the American Psychological Association (2013).

Limitations of the Research Design

Causal comparative design is a nonexperimental type of quantitative research. Manipulation or control of variables is not used, and application of an intervention has already occurred. Causal comparative design can describe relationships, but causation cannot be ensured (Adams & Lawrence, 2014; Creswell, 2014).

Some limitations of the study were inherent to the use of existing gradebooks. The collected data did not include recorded NCLEX-RN pass results for all students. NCLEX-RN pass results were tracked for only 40 students, including those students with individual test averages of less than 75% but whose final averages including collaborative testing points were 75% or higher. Collaborative testing points allowed these students to pass a course they would have failed with only individual testing, making their NCLEX-RN results of particular interest to faculty.

Sampling is another limitation of a nonexperimental study using existing data. Random sampling was not used, increasing the chance that the study sample does not well represent the full population of nursing students. Instead, nonprobability sampling was used, with an increased chance for sampling bias (Adams & Lawrence, 2014). Data represents students from a nursing program serving a larger number of minority students than is found in most locations.

Collaborative testing was used as a second test in this study. Collaborative testing can be used in numerous ways, but this study was based on a single approach. In addition, students in this study were exposed to a variety of collaborative learning experiences, including collaborative case studies, group projects, presentations, and concept mapping. Active and team learning are uncontrolled variables that can also influence student performance.

This was a causal comparative study and does not imply that collaborative testing was the cause of student success. Uncontrolled variables could have contributed to successful program completion and student success on the NCLEX-RN, including, but not limited to, faculty teaching skill, additional student support services, use of NCLEX-RN preparation programs, and student motivation. Conversely, other uncontrolled variables could have contributed to program or NCLEX-RN failure, including, but not limited to exposure to inexperienced or unskilled faculty, time and life demands, illness, and lack of student motivation.

Internal and External Validity

There was no manipulated independent variable, or control of other variables, in this causal comparative (ex post facto) design. Use of existing data prohibited control of variables and presented increased threat to internal validity. Effect size was then used as an indicator of internal validity and reliability, especially because validated measures, tools, or inventories were not used (Adams & Lawrence, 2014). Student performance in this study was directly assessed using percentage correct on course exams. External validity, or generalizability, is limited by sample size and composition. The sample for this study represented a higher proportion of minority students than is found in most schools of nursing.

Expected Findings

I expected that these data would show collaborative testing positively correlates with course success. However, the strength of this relationship was unknown, along with differences in the benefits of collaborative testing experienced by the top performers in the sample compared with the lower performers in the sample, with grouping based on the 75% final test average cutoff required to pass the course. Some students passed this advanced medical-surgical nursing course by earning collaborative testing points. These students would have failed the course

without collaborative testing points, and possibly failed out of the nursing program, but instead were able to successfully complete the course and progress in the program. An additional unknown was whether these students that benefited from passing the course with additional collaborative testing points went on to complete the program and pass the NCLEX-RN on first attempt. If collaborative testing were to be found to strongly correlate with student success and stable NCLEX-RN pass rates, faculty may increase use of testing for learning with a diminished fear of negatively impacting program outcomes.

Ethical Issues in the Study

Causal comparative design limits ethical considerations because there is no manipulation of variables by the researcher. Instead, application of an intervention is examined retrospectively to identify patterns and relationships that are occurring naturally (Adams & Lawrence, 2014). In this study, data represents test scores before and after application of an educational approach that benefitted student grades. This intervention was applied equally to all students, and all students, both lower and higher performers, had the same chance to benefit both in learning and improved course grades. Students were not at risk to experience any negative effects from the application of the intervention. These data used in this study were extracted from seven semesters of gradebooks and had been de-identified by the instructor prior to receipt of permission from the institution and inclusion in this study.

Potential ethical issues identified in the Belmont Report that became the basis for the Federal Common Rule are respect for persons, beneficence, and justice (Adams & Lawrence, 2014). These data used in this study were reviewed in light of ethical application of the intervention. The application of collaborative testing in this study was applied to all students represented by the pre-existing gradebook data, and all had an equal chance to benefit. Test

scores within the data were increased by the intervention, which led to an increase in the test score average on which the course grade was based. The data shows that nine students over seven semesters passed the advanced medical-surgical nursing course that they would have failed based on individual test scores alone. There were no grades negatively impacted by the collaborative testing points represented by this data. These facts demonstrate an application that adheres to the principles of respect for persons, beneficence, and justice that is required for research.

The university IRB representative was contacted for permission to use previously collected gradebook data for this research. An email response stated, “De-identified data sets do not meet the definition of human subjects and thus does not require IRB review” (see Appendix A). The IRB at the institution in the southwest thus declined to review a research proposal. All Concordia University-Portland IRB protocol were followed, and permission to use the de-identified data was granted in March 2018.

Summary

In this chapter I explained the causal comparative design I implemented to examine relationship between collaborative testing and student success. Additionally, differences in this relationship were explored by comparing higher and lower achieving students. Finally, the data were used to determine if there was a relationship between course success due to collaborative testing, and first-time NCLEX-RN pass rates.

Collaborative testing uses the principles of adult learning, combining assessment with a continued opportunity for students to better understand applied knowledge. Learning is enhanced by application of knowledge to real-life situations. Based on adult and transformational learning

theories, this approach to applied knowledge combined with collaboration and reflection, is effective for the goal-oriented adult (Knowles et al., 2015; Merriam, 2001).

Because schools of nursing rely heavily on NCLEX-RN style testing to evaluate student learning, it is important to optimize that time and better prepare students for practice. However, nursing faculty have resisted deviating from a traditional approach to testing that emphasizes a controlled and secure testing environment that measures only individual performance. This is at least partly the result of a need to protect program NCLEX-RN first-time pass rates. A better understanding of how application of the constructivist principles of adult learning correlates with end-of-program outcomes can allow nursing faculty to adopt proven educational strategies while still protecting program NCLEX-RN pass rates.

Chapter 4: Results

Introduction

The purpose of this study was to analyze the traditional use of testing in schools of nursing by examining the relationship between use of nontraditional collaborative testing and nursing student success. Collaborative testing has previously been shown to increase learning and critical thinking skills, increase communication and collaboration, and decrease test anxiety (Duane & Satre, 2014; Martin et al., 2014; Peck et al., 2013; Sandahl, 2009; Vogler & Robinson, 2016, 2016; Zhang & Henderson, 2017) However, research on the relationship between collaborative testing and overall course success has not been thoroughly examined in schools of nursing where heavy use of traditional and high-stakes testing are the norm (Molsbee, 2013). In this ex-post facto study, I used seven semesters of archival gradebook data wherein both individual and collaborative test scores were available for analysis. The study was delimited by the use of existing de-identified data, and no new data were collected.

This chapter provides the results. Presented first is a description of the sample and a recap of the data collection procedure, followed by the data analysis. Three research questions were of primary interest and are listed with their associated null and alternative hypotheses in the data analysis section. The chapter concludes with a short summary.

Description of the Sample

Test score data were used that represented seven consecutive semesters of nursing students ($N = 131$) in an advanced medical-surgical course. Class sizes ranged from 13 to 25 students each semester. Data collection took place between August 2008 and December 2011 on a small campus of a major state university in the southwestern region of the United States.

Data included six individual tests each semester. Each test consisted of 50 to 100 multiple-choice or multiple-select style questions. Multiple select questions allow students to choose multiple answers or select all that apply. After each individual test, the student was randomly assigned to a collaborative test group by drawing a number from a bowl. Group sizes ranged from four to six students. Each group then took the same test again as a group. Thus, data included two scores for each test—an individual score and a collaborative score. The collaborative score was obtained by adding percentage points to the individual score based on the group's performance. For groups who scored 90%–99% correct, 2% was added to the individual score, which became the collaborative score. For groups who scored 100% correct, 5% was added to the individual score, which became the collaborative score. Groups who scored below 90% received no addition of points to their individual scores. Although all students participated in the collaborative testing groups, they had to have scored 75% or greater on the particular individual test to qualify for collaboration points. The 2% and 5% adjustments to the individual score were based on the idea that it would provide incentive for participation without greatly inflating test scores. The collaborative testing procedures and point adjustments described above were established as a result of a conversation with other nurse educators at a professional nurse educator conference. (More detail was provided in Chapter 3.)

Data comprised six individual scores, six collaborative scores, and the total number of collaborative testing points earned over the semester for each student represented. The data set comprised the de-identified final gradebooks over the seven semesters. In addition, NCLEX-RN pass/fail data were available for 40 of the 131 entries. The gradebook data were transferred into SPSS Version 25 for the subsequent statistical analyses.

Summary of the Results

Pearson's correlation coefficient was chosen as the statistical test to explore RQ1. This statistical test provides information on direction and strength of the relationship between two variables. Analysis showed a strong positive relationship between collaborative testing points and final test averages. An independent-sample *t* test was used for RQ2, to compare the collaborative testing points earned for those with final test averages above and below the course success benchmark of .75. Cohen's *d* was calculated to determine the magnitude of the identified differences (Field, 2005). These results led to inclusion of an additional *t* test, examining differences between the two groups' final test averages. For RQ3, chi-square was used to analyze NCLEX-RN pass rates for the 40 students with available results. No further modifications were indicated during data analysis.

All data were screened for missing data and outliers. Existing gradebooks contained complete records of test scores, individual test averages, and final averages with the addition of collaborative testing points used for RQ1 and RQ2. No outliers were found that greatly deviated from other scores, or that would threaten to skew the data. All 40 student NCLEX-RN pass results were used to compare groups using chi-square analysis for RQ3.

Detailed Analysis

Each research question is listed followed by its associated statistical analysis. The .05 level of probability was the criterion used for testing the null hypotheses. For analysis purposes, the individual scores and collaborative scores were averaged, giving each student an individual test average and an average after addition of collaborative testing points. The overall average collaborative score and total collaborative testing points earned were the basis for the statistical

analyses. There were no missing data for collaborative testing points, individual test scores, or final test scores with collaborative testing points added.

RQ1. What is the relationship between collaborative testing points earned and student success based on final course test averages in an advanced medical-surgical nursing course?

H₀. There is no relationship between earned collaborative testing points and final course test averages in a medical-surgical nursing course.

H_a. There is a relationship between earned collaborative testing points and final course test averages in a medical-surgical nursing course.

Pearson correlation was used to answer RQ1. As part of the procedure, the scores were screened for extreme scores, high or low, that might influence the results, as well as for the assumptions of linearity and normality. No outliers were identified and the assumptions were not violated.

Table 2 provides the results. The correlation was positive and statistically significant ($r = .82, p < .01$) indicating that as the test averages tended to increase the collaborative testing points also tended to increase. Thus, the null hypothesis was rejected, showing support for the alternative hypothesis.

Although the correlation is significant, statistical significance provides no information about how strong or important a correlation may be. Rather, it only provides the probability (p) of observing results as extreme as those that would support the null hypothesis. For this reason effect sizes be reported in conjunction with statistical significance results (American Psychological Association, 2013). Effect size is an indicator of the strength of a correlation regardless of whether it is statistically significant or not. For correlation, the correlation coefficient itself can be interpreted as an effect size. Cohen (as cited in Field, 2005) suggested

interpreting the coefficient as an effect size where $r = .10$ (small effect); $r = .30$ (moderate effect); $r = .50$ (strong effect). Thus, from an effect size perspective, the correlation of .82 indicates a strong relationship between collaborative testing points earned and the final test averages (Field, 2005).

Table 2

Intercorrelation for Collaborative Testing (CT) Points Earned and Student Final Test Average
($N = 131$)

Measure	<i>M</i>	<i>SD</i>	<i>r</i>
CT points earned	8.88	6.33	.82**
Final test average	.81	.07	

** $p < .01$

RQ2. What is the difference in the total collaborative testing points earned between the lower performing students and higher performing students from a medical-surgical nursing course?

H_0 . There is no difference in the average collaborative testing points earned between the lower and higher performing students.

H_a . There is a difference in the average collaborative testing points earned between the lower and higher performing students.

The low performers were defined as those who failed the course where the final test average was below 75% ($n = 22$). The high performers were defined as those whose final test average was 75% or above ($n = 109$). An independent samples t test was used to answer RQ2. The independent variable was pass or fail, and the dependent variable was the collaborative

testing points. Prior to conducting the analysis, the groups were screened for outliers and none were identified.

Row 1 of Table 3 provides the results for RQ2. Observation of the means shows that the collaborative testing points for the higher performers was greater than the lower performers ($M = 9.98$ vs. $M = 3.41$, respectively). While the assumption of normality was met, the assumption that the variances were equal was not met (Leven's test $p < .05$); the SD s were considerably different ($SD = 6.27$ versus $SD = 2.81$). Taking this into consideration, and not assuming equal variances, the difference of 6.57 points between the means was statistically significant ($t = 7.75$, $p < .000$). Thus, the null hypothesis was rejected and the alternative hypothesis accepted with the group that passed showing a higher number of points earned.

As with correlation, statistical significance between means provides no evidence about how important or large a difference may be. The last column (Cohen's d) shows the effect size and is an indicator of how strong or important the difference may be, where $d = .20$ (small difference); $d = .50$ (moderate difference); and $d = .80$ (large difference). Using these rules (Field, 2005), the difference between the two groups can be considered large ($d = 1.35$).

The second row of Table 3 provides a supplementary t test analysis between the final test averages of the two groups. As expected, the group that passed had a higher test average of .83 than those who did not pass, .72, and that the difference of .11 was statistically significant ($t = 10.77$, $p < .001$). From an effect size perspective, using the above criteria, the difference can be considered large ($d = 1.93$), which suggests that the use of collaborative testing did not overly inflate passing rates for the lower performing students.

RQ3. What is the difference in the overall NCLEX-RN first-time pass rates and the NCLEX-RN first-time pass rates of those who passed the course due to collaborative testing?

H₀. NCLEX-RN pass rates for students who pass an advanced medical-surgical nursing course due to collaborative testing points are not significantly different than those of their graduating cohorts.

H_a. NCLEX-RN pass rates for students who pass an advanced medical-surgical nursing course due to collaborative testing points is significantly different than those of their graduating cohorts.

Table 3

Group Differences on CT Points, and Final Test Average

Measures	Passed Course		Failed Course		Diff	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
CT Points	9.98	6.27	3.41	2.81	6.57	7.75	.00***	1.35
FTA	.83	.07	.72	.04	.11	10.77	.00***	1.93

Note. FTA = Final test average; Diff = the difference between the two means.

*** $p < .001$

A chi-square test for independence with Yate's continuity correction was used to test for an association between the NCLEX-RN pass rates of those who passed the course due to addition of collaborative testing points and those who did not. Yate's continuity correction is used for a 2 x 2 chi-square analysis, which tends to produce small significance values, increasing the risk of a type I error (Field, 2005).

Table 4 provides the results that indicated no statistically significant association between course success due to collaborative testing points and passing the NCLEX-RN on first attempt ($N = 40$, $\chi^2 = .023$, $p = .88$). Thus, the null hypothesis was retained, indicating that there was no

significant difference in NCLEX-RN pass rates for those who passed the course because of collaborative testing points and those who would have passed without the additional points.

For effect size, the phi correlation coefficient was obtained as part of the chi-square analysis showing the correlation between the pass rates and pass/fail (phi = .14). The phi coefficient is equivalent to the Pearson correlation coefficient, and the same criteria may be used to indicate the strength of the relationship where phi = .10 (small effect); phi = .30 (moderate effect); phi = .50 (strong effect, Adams & Lawrence, 2014). Using these criteria for effect, the relationship between the pass rates and pass/fail can be considered between small and moderate.

Table 4

NCLEX-RN First Attempt Pass Rates for Students With Course Pass/Fail Status Change due to Collaborative Testing Points

	NCLEX-RN Pass		NCLEX-RN Fail		χ^2	<i>p</i>	phi
	<i>n</i>	%	<i>n</i>	%			
Course status change	8	100	0	0	.023	.88	.14
No status change	29	90.6	3	9.4			

Chapter Summary

Results showed a strong positive relationship between collaborative testing points and final course test averages. Comparison of group means for lower and higher performing students revealed a statistically significant difference and large effect size, with higher performing students earning more collaborative testing points than lower performing students. There was also a statistically significant difference between groups on final course average, with an effect

size that indicated the difference was large. There was no difference in NCLEX-RN first-time pass rates for those who passed the course due to the addition of collaborative testing points, and those who would have passed without that addition. These results will be discussed in Chapter 5.

Chapter 5: Discussion and Conclusion

Researchers have inadequately examined the relationship between nursing education practice and the use of testing. Nursing education practice favors traditional testing of knowledge retention (Halstead, 2013; Kantar, 2014; Killingsworth, Kimble, & Sudia, 2015; Oermann & Gaberson, 2017), while research demonstrates the value of testing for continued learning (Foss & Pirozzolo, 2017; Griswold et al., 2017; Harrison & Wass, 2016; Mbalamula, 2018; Racsmány et al., 2018; Raupach et al., 2016; Yang & Shanks, 2018). Collaborative testing is an approach that has many benefits, including increased learning, decreased test anxiety, and development of collaboration skills (Baghdady et al., 2014; Hanna, et al., 2016; Peck et al. 2013; Raupach et al., 2016; Siegel et al., 2015; Vogler & Robinson, 2016; Zhang & Henderson, 2017). Collaborative testing allows groups of students to work together on a test, coming to consensus on the best answer for each question. For this study, collaborative testing was used following individual testing as a second test, with points added to individual scores based on the collaborative test score. This study added to the body of knowledge regarding creative use of testing for continued learning that takes advantage of testing effect, extension of critical thinking and inductive reasoning, and collaborative problem-solving. The relationship between use of collaborative testing and student success was explored to better determine the value of collaborative testing.

In this chapter I discuss results of the statistical analysis of seven semesters of gradebook data ($N = 131$). I interpret the findings discuss how this study has added to the relevant literature. Limitations will be identified, along with implications for practice, and recommendations for future research.

Summary of the Results

Adult and transformational learning theories call for use of active learning strategies with adults who are motivated to learn (Merriam, 2001). Research has suggested integrating testing and learning uses the benefits of testing effect to increase retention of learning, improve critical thinking and students' ability to generalize knowledge, and use inductive reasoning to problem solve questions they have not previously encountered (Griswold et al., 2017; Iwamoto et al., 2017; Racsmány et al., 2018; Yang & Shanks, 2018). Distraction can decrease learning during traditional studying but has little impact on learning through testing (Buchin & Mulligan, 2017). Increased frequency of course testing has demonstrated a positive impact on final exam scores (Foss & Pirozzolo, 2017). Students perform better on collaborative tests than individually, with random groups outperforming student selected groups (Mbalamula, 2018). Requiring students to provide answer rationales, which is part of group discussion when reaching answer consensus with collaborative testing, leads to better long-term retention and deeper learning (Zhang & Henderson, 2017). Collaborative testing as a second test leads to greater retention of learning than individual testing with answer feedback (Vogler & Robinson, 2016). Continued support of testing for learning in the literature demonstrates that there are many ways to integrate testing into learning, all with positive impact on student learning outcomes. The results of this study are intended to provide information related to barriers that limit adoption of collaborative testing in nursing education. Results are listed according to research question.

RQ1. What is the relationship between collaborative testing points earned and student success based on final course test averages in an advanced medical-surgical nursing course?

Results supported rejection of the null hypothesis and acceptance of the alternative hypothesis, indicating a strong relationship between earned collaborative testing points and end-

of-course test averages. Student success is an important goal for nurse educators, and preventing course failure positively affects retention and completion rates. Retention and completion rates are potential indicators of program quality, along with NCLEX-RN pass rates (Gibson, 2014; Halstead, 2013).

RQ2. What is the difference in the total collaborative testing points earned between the lower performing students and higher performing students from a medical-surgical nursing course?

Students were divided into two groups, using the course passing benchmark of .75 to separate students into the higher performing and lower performing groups. Comparison of groups revealed that the higher performers earned more collaborative testing points than lower performers. Results related to RQ2 indicated that lower performing students do not disproportionately benefit from the rewarding of collaborative testing points. Rather, the high performers' test averages benefited the most from collaborative testing points. The null hypothesis was rejected in favor of the alternative hypothesis, positing a difference between the collaborative testing points earned between the lower and higher performing students. Additionally, comparison of groups based on final test averages indicated a significant difference, and large effect. This suggests that pass rates were not overly influenced by the addition of collaborative testing points, and lower performers' test scores were not disproportionately inflated through the use of collaborative testing with additional points awarded.

RQ3. What is the difference in the overall School of Nursing NCLEX-RN first-time pass rates and the NCLEX-RN first-time pass rates of those who passed a course due to collaborative testing?

The preexisting gradebook data did not record NLCEX-RN pass rates for all students. Forty results were recorded, which represented just over 30% of students. I compared NCLEX-RN success for students who passed the course because of collaborative testing points and those who would have passed the course without collaborative testing points. Results indicated that there was no difference in NCLEX-RN pass rates, and the null hypothesis was retained. This result supports nurse educators should use collaborative testing without fear of a decline in program NLCEX-RN pass rates.

Discussion of the Results

In this causal comparative study, I examined the relationship between collaborative testing and student success. Results provided additional support for the use of this testing approach in schools of nursing. In spite of the learning benefits of testing found in the literature because of testing effect and of collaborative testing in particular, nurse educators continue to conduct nursing courses in a traditional way based on teaching content and testing individually for retention. This traditional approach fails to promote student development of higher-order reasoning skills. “There is an urgent need for transforming educators’ beliefs, knowledge, and skills on testing” (Kantar, 2014, p. 789). Collaborative testing provides continued learning during testing and is easily incorporated into existing courses with little additional time, energy, or expense. Nurses currently enter practice by stepping into positions of increasing responsibility that require well developed teamwork, collaboration, critical thinking, and leadership skills (American Association of Colleges of Nursing, 2014; Institute of Medicine, 2010). Collaborative testing has demonstrated potential for helping to develop these 21st-century skills (Duane & Satre, 2014; Martin, Friesen, & De Pau, 2014). Concern about the impact of nontraditional testing on student success has been a barrier to change for nurse educators. In this study,

quantitative analysis of existing gradebook data provided findings to support a change, which is further discussed in this section.

Results of Pearson correlation used for RQ1 showed a strong positive relationship between collaborative testing points and student success, as measured by end-of-course test averages. This was an expected finding, since collaborative testing points were added to the student's individual exam score to produce the combined score used to calculate student grades. Although it was possible for students to earn no collaborative testing points, it was not possible for the collaborative test to result in a decrease in student test scores.

A comparison of lower and higher performing groups for RQ2 found that earned collaborative testing points increased as student scores increased. In other words, students who were high scorers tended to earn more collaborative testing points in spite of mixed testing groups and equal opportunity for all students to earn the same number of collaborative points. Nurse educators are particularly concerned about possible grade inflation, and enabling under-prepared, low scoring students to pass a course. Attempts to address this concern led to the additional comparison of the final test averages of the two groups. Although there was a significant difference found in final test averages for the lower and higher performing students, with a large effect size, results indicated collaborative testing points did not greatly influence final test averages for lower performers and lead to grade compression.

Nurse educators share a concern that students who pass a course due to additional points earned through collaborative testing will graduate unable to pass the NCLEX-RN. Addressing this concern was the objective of RQ3, which tested for an association between overall NCLEX-RN pass rates, and the NCLEX-RN pass rates of those who passed the course due to collaborative testing points. The data did not include NCLEX-RN results for all students;

however, just more than 30% of student results were available, including all students who passed the course because of collaborative testing points. There was no statistically significant association found between NCLEX-RN success and passing the course irrespective of collaborative testing points. The results showed that those who passed the course after collaborative testing who would have failed without it demonstrated NCLEX-RN success at a rate equivalent to other students.

Discussion of the Results in Relation to the Literature

The literature shows that collaborative testing has previously demonstrated benefits for student learning, critical thinking, decreased anxiety, and increased collaboration skills (Baghdady et al., 2014; Hanna et al., 2016; Peck et al., 2013; Raupach et al., 2016; Siegel et al., 2015; Vogler & Robinson, 2016; Zhang & Henderson, 2017). My study has attempted to show whether the use of collaborative testing positively correlates with student success in schools of nursing, where testing is currently used for individual assessment of learning, and as a high-stakes measure to make progression and graduation decisions (Halstead, 2013; National League for Nursing, 2012; Oermann & Gaberson, 2017). A more student-centered approach, grounded in constructivism and the principles of adult and transformational learning theories, would require a culture change in nursing education, where testing helps to prevent graduation of students who might not pass the NCLEX-RN, and thus threaten program accreditation and state Board of Nursing approval status (Halstead, 2013; National League for Nursing, 2012). Evidence that collaborative testing can be used for continuation of learning without enabling weak students to jeopardize program status would go far in easing the fears of nurse educators.

Statistical analysis related to RQ1 demonstrated a strong positive relationship between collaborative testing points and end-of-course test averages. This finding indicates that

collaborative testing can positively contribute to course success. However, this contribution to test scores could be viewed as grade inflation, allowing weaker students to pass a course they could not pass working alone (Jang et al., 2017). Comparison of lower and higher performers for RQ2 was performed in order to examine whether weaker students disproportionately benefited from collaborative testing. However, the opposite was found to be true. Earned collaborative testing points of higher performers were greater than those of lower performers. Additionally, the effect collaborative testing points had on final exam averages was small and did not produce grade inflation.

Nursing literature confirms that nurse educators use tests to prevent progression and graduation of low performers in order to protect program NCLEX-RN first time pass rates (Halstead, 2013; National League for Nursing, 2012). Educators are wary of approaches that could artificially inflate the grades of low scorers, allowing them to pass a course in which they have not mastered the content (Centrella-Nigro, 2012; Donaldson & Gray, 2012; Duane & Satre, 2014). Professional nursing organizations warn against this use, cautioning nurse educators to use multiple approaches to evaluate student knowledge and skills (Halstead, 2013; Kantar, 2014; National League for Nursing, 2012; Rizzolo, 2015). Additionally, nurse educators have an objective to increase the number of men and underrepresented minorities entering the profession, which has traditionally been dominated by White females (Ferrell et al., 2016). Educational approaches that support student success are especially important for students who have not traditionally done well in higher education: first-generation college students, English as a second language students, and other underrepresented groups (Eastridge, 2014; Ferrell et al., 2016).

Nursing educators have recently begun to move away from traditional testing and adopt more creative approaches. The National League for Nursing (NLN; 2012) began a project that

explores use of simulation to assess nursing students' end-of-program knowledge and clinical reasoning skills in the areas of patient assessment and intervention, clinical judgment, safety, and collaboration. One lesson learned through this project was the complexity involved in ensuring consistent evaluation of student performance during simulation testing. Inter-rater reliability for simulation evaluation has been difficult to achieve. Other conclusions from this project corresponded with the NLN Fair Testing Guidelines (NLN, 2012), suggesting no student should be evaluated based on a single test. However, because use of simulation for testing is costly and time intensive, it is an impractical solution for widespread adoption. Moreover, this project still centers around use of testing for evaluation of learning, ignoring the potential use of testing for continued learning. Additional options for testing that support continued learning need to be available.

Although benefits of collaborative testing have been uncovered through educational research (Baghdady et al., 2014; Hanna et al., 2016; Peck et al., 2013; Raupach et al., 2016; Siegel et al., 2015; Vogler & Robinson, 2016; Zhang & Henderson, 2017), the relationship between use of collaborative testing and nursing student success, including NCLEX-RN results, has not been previously well addressed, and is one barrier to broad adoption of this learning tool. One study (Molsbee, 2013) addressed this gap that prevents application of research to practice and tracked students who passed courses due to collaborative testing in order to find if they went on to complete the nursing program and pass the NCLEX-RN. Molsbee (2013) found that students who passed one course based on collaborative testing were likely to complete, while those who passed two or more courses due to collaborative testing had less chance of completing and passing the NCLEX-RN. Although these findings addressed the broader issue of student long-term success related to collaborative testing, no statistical analysis were done to determine

significance or power of these observations. Nurse education researchers are called upon to address identified gaps in the literature. My study adds to current knowledge through statistical analysis of existing gradebooks for a course that consistently used collaborative testing. This study adds to knowledge about the relationship between use of collaborative testing and student success, including the knowledge gap related to NLCEX-RN success for students who benefit from collaborative testing. The data demonstrated that students who derive grade benefits from collaborative testing, passing a course they would not have passed without it, go on to pass the NCLEX-RN with a success rate equal to their peers.

Limitations

This study examined one use of collaborative testing: collaborative testing as a second test with points added to students' individual test scores based on collaborative test scores. Testing can be employed creatively in many ways so that students benefit from the testing effect. Collaborative testing can be used as a primary test, a pretest, a second test, or as part of a formative quizzing process. Points can be awarded for collaborative testing based on different criteria, or collaborative scores may be averaged with individual scores (Heglund & Wink, 2011). Data representing different applications of collaborative testing would enhance existing research.

Using existing gradebook data did not allow me to control for variables. Students may have experienced additional interventions that impacted student success. While my research examined relationship, causation cannot be determined. Student demographic data was not available with the existing gradebook data, and NCLEX-RN results were available for just over 30% of the data, which limits the conclusions that can be drawn based on existing analysis.

This study was based on gradebook data for students taking a course during the final semester of a nursing program that did not widely use collaborative testing in other courses. Thus, I did not consider passing multiple courses, which made it difficult to compare the findings to the one other study found related to collaborative testing and nursing student success (Molsbee, 2013).

Implication of the Results for Practice, Policy, and Theory

Adult and transformational learning theories call for using active learning strategies with adults who are motivated to learn (Merriam, 2001). Nurse educators are charged with preparing nursing graduates for modern practice, which requires the nurse to work collaboratively in caring for increasingly complex patients. A traditional approach of teaching facts followed by testing for recall is no longer adequate (American Association of Colleges of Nursing, 2014; Institute of Medicine, 2010; Kantar, 2014). Collaborative testing is one way to better prepare graduates for practice while decreasing student attrition by increasing retention rates for borderline passing students. This research demonstrated a positive relationship between use of collaborative testing and course success, with little effect on grade inflation. Collaborative testing points did not overly benefit lower scoring students. In addition, students that succeeded due to collaborative testing were shown to successfully pass the NCLEX-RN after graduation.

Nurse educators are cautioned to balance use of high-stakes testing with low-stakes, low-anxiety approaches (National League for Nursing, 2012; Røykenes, Smith, & Larsen, 2014; Spurlock, 2013). Minority students are often under-prepared for the academic rigor of higher education, and have higher attrition rates (Ferrell et al., 2016; Jefferys, Hodges, & Trueman, 2017). Collaborative testing may benefit these at-risk students through improved test scores, retention, and graduation rates.

It is important that educators identify approaches to learning that have little or no impact on budgets, and require minimal faculty training time. Creative use of testing does not consume limited resources, and requires minimal time commitment. In addition, students are not required to purchase additional books or products from textbook or testing companies. Cost and energy demands are small, and collaborative testing is easy to incorporate into existing curriculum, making it an ideal approach for immediate use in the classroom.

Recommendations for Further Research

The existing gradebook data did not include NCLEX-RN results for all represented students. Replicating this study with complete data and the addition of student demographic data could allow for additional comparisons and other conclusions. In addition, this study was based on one application of collaborative testing as a second test. Further research is needed to compare applications of collaborative testing as a primary test or as a pretest.

Recommendations that are clear in the literature include random group assignment that results in better learning than student selected groups; potential grade benefits that, no matter how small, encourages student enthusiasm and participation; and frequent testing that produces better long-term retention. How this knowledge is used is a matter for additional research.

Applied research is needed that measures learning through collaborative testing based on cognitive level and difficulty of the learned material. Students process knowledge-based learning and memorization through different learning pathways than complex reasoning, which requires consideration of multiple factors to solve a clinical problem. Studies have shown that collaborative testing improves learning overall, but results are mixed based on the difficulty of the subject and long-term retention. Demographic differences between students also merit attention, so that educators can better identify measures that promote learning for diverse

classrooms, and promote success for students who are historically more likely to fail in higher education.

Additionally, end-of-program outcomes for students who pass more than one course due to the benefits of collaborative testing should be revisited in light of the observations reported by Molsbee (2013), who reported that students who passed more than one course due to collaborative testing were more at risk for attrition and NCLEX-RN failure.

Conclusion

Testing is a powerful learning tool due to testing effect, and collaborative testing offers additional benefits that prepare nursing students for the challenges of the modern workplace. Nurse educators currently rely on traditional and high-stakes testing for evaluation of learning. Expanded use of collaborative testing for continued learning offers additional benefits that include practice with communication and group process, decreased test anxiety, self-reflection and analysis of thinking, and diminished negative testing effects.

Collaborative testing has frequently been researched through use of student and faculty surveys, which provide information about attitudes and perceptions. This study has used gradebook data which reveal that collaborative testing correlates positively with student success without disproportionately inflating the grades of lower performing students. In addition, students who benefited from collaborative testing points, passing a course they would have otherwise failed, demonstrated an ability to go on and complete the nursing program and pass the NCLEX-RN on first attempt. Analysis of gradebook data may begin to relieve the fears of nurse educators who are hesitant to deviate from a traditional model of individual and high-stakes testing in order to protect program NCLEX-RN pass rates.

Schools of nursing are currently in a state of transition as nurse educators prepare graduates to work in an increasingly complex health care environment. Increasing demand for critical thinkers who can work collaboratively to solve complex problems has led to a need for a culture change from an assessment of learning culture to an assessment for learning culture, one that aligns with educational research on best testing practices and better supports student success. This study adds to the body of knowledge that can move nurse educators forward toward accomplishing this change.

References

- Abel, M., & Roediger, H. L. (2017). Comparing the testing effect under blocked and mixed practice: The mnemonic benefits of retrieval practice are not affected by practice format. *Memory & Cognition*, *45*(1), 81–92. <https://doi.org/10.3758/s13421-016-0641-8>
- Adams, K. A., & Lawrence, E. K. (2014). *Research methods, statistics, and applications*. Thousand Oaks, CA: Sage.
- American Association of Colleges of Nursing. (2014). *Fact sheet: The impact of education on nursing practice*. Retrieved from <http://www.aacn.nche.edu/media-relations/EdImpact.pdf>
- American Psychological Association. (2013). *Publication manual of the American Psychological Association* (6th ed.). Washington DC: American Psychological Association.
- Atabek Yigit, E., Balkan Kiyici, F., & Çetinkaya, G. (2014). Evaluating the testing effect in the classroom: An effective way to retrieve learned information. *Eurasian Journal of Educational Research*, *54*, 99–116.
- Baghdady, M., Carnahan, H., Lam, E. W. N., & Woods, N. N. (2014). Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Medical Education*, *48*(2), 181–188. <https://doi.org/10.1111/medu.12302>
- Bailey, P. H., Mossey, S., Moroso, S., Cloutier, J. D., & Love, A. (2012). Implications of multiple-choice testing in nursing education. *Nurse Education Today*, *32*(6), e40–e44. <https://doi.org/10.1016/j.nedt.2011.09.011>
- Barton, L., Willson, P., Langford, R., & Schreiner, B. (2014). Standardized predictive testing: Practices, policies, and outcomes. *Administrative Issues Journal: Connecting Education, Practice, and Research*, *4*(2). <https://doi.org/10.5929/2014.4.2.2>

- Bennett, R. E. (1998). *Reinventing assessment: Speculations on the future of large-scale educational testing. A policy information perspective.* (No. ED 424-254). Princeton, NJ: United States Department of Education, Policy Information Center, Educational Testing Service. Retrieved from <https://eric.ed.gov/?id=ED424254>
- Berndt, J., Dinndorf-Hogenson, G., Herheim, R., Hoover, C., Lang, N., Neuwirth, J., & Tollefson, B. (2015). Collaborative classroom simulation (CCS): An innovative pedagogy using simulation in nursing education. *Nursing Education Perspectives, 36*(6), 401–402. <https://doi.org/10.5480/14-1420>
- Bines, J. E., & Jamieson, P. (2013). Designing new collaborative learning spaces in clinical environments: Experiences from a children’s hospital in Australia. *Journal of Interprofessional Care, 27*, 63–68. <https://doi.org/10.3109/13561820.2013.795933>
- Bovee, B. (2016). The impact of collaborative testing on test anxiety. *Chiropractic Journal of Australia, 44*(3), 215–221.
- Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language, 66*(3), 407–415. <https://doi.org/10.1016/j.jml.2011.12.009>
- Buchin, Z. L., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1934–1947. <http://dx.doi.org/10.1037/xlm0000427>
- Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory & Cognition, 43*(2), 193–205. <https://doi.org/10.3758/s13421-014-0462-6>

- Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, 21(5), 279–283. <https://doi.org/10.1177/0963721412452728>
- Centrella-Nigro, A. M. (2012). Collaborative testing as posttest review. *Nursing Education Perspectives*, 33(5), 340–341.
- Chen, H.-Y., & Chuang, C.-H. (2012). The learning effectiveness of nursing students using online testing as an assistant tool: A cluster randomized controlled trial. *Nurse Education Today*, 32(3), 208–213. <https://doi.org/10.1016/j.nedt.2011.03.004>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.
- Dahlstrom, O. (2012). Learning during a collaborative final exam. *Educational Research and Evaluation*, 18(4), 321–332. <https://doi.org/10.1080/13803611.2012.674689>
- Donaldson, J. H., & Gray, M. (2012). Systematic review of grading practice: Is there evidence of grade inflation? *Nurse Education in Practice; Kidlington*, 12(2), 101–114. <http://dx.doi.org.cupdx.idm.oclc.org/10.1016/j.nepr.2011.10.007>
- Duane, B. T., & Satre, M. E. (2014). Utilizing constructivism learning theory in collaborative testing as a creative strategy to promote essential nursing skills. *Nurse Education Today*, 34(1), 31–34. <https://doi.org/10.1016/j.nedt.2013.03.005>
- Eastridge, J. A. (2014). Use of collaborative testing to promote nursing student success. *Nurse Educator*, 39(1), 4–5. <https://doi.org/10.1097/01.NNE.0000437366.96218.f1>
- Fernández Alemán, J. L., Carrillo de Gea, J. M., & Rodríguez Mondéjar, J. J. (2011). Effects of competitive computer-assisted learning versus conventional teaching methods on the acquisition and retention of knowledge in medical surgical nursing students. *Nurse Education Today*, 31(8), 866–871. <https://doi.org/10.1016/j.nedt.2010.12.026>

- Ferrell, D. K., Decrane, S. K., Edwards, N. E., Foli, K. J., & Tennant, K. F. (2016). Minority undergraduate nursing student success. *Journal of Cultural Diversity, 23*(1), 3–11.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Foss, D. J., & Pirozzolo, J. W. (2017). Four semesters investigating frequency of testing, the testing effect, and transfer of training. *Journal of Educational Psychology, 109*(8), 1067–1083. <http://dx.doi.org/10.1037/edu0000197>
- Freda, N. M., & Lipp, M. J. (2016). Test-enhanced learning in competence-based predoctoral orthodontics: A four-year study. *Journal of Dental Education, 80*(3), 348–354.
- Gibson, H. A. (2014). A conceptual view of test anxiety. *Nursing Forum, 49*, 267–277.
- Gilley, B. H., & Clarkston, B. (2014). Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching, 43*(3), 83–91.
- Gordon, E. W., & Pajagopalan, K. (2016). *The testing and learning revolution: The future of assessment in education*. New York: Palgrave Macmillan.
- Gorvine, B. J., & Smith, H. D. (2015). Predicting student success in a psychological statistics course emphasizing collaborative learning. *Teaching of Psychology, 42*(1), 56–59. <https://doi.org/10.1177/0098628314562679>
- Griswold, L. A., Overson, C. E., & Benassi, V. A. (2017). Embedding questions during online lecture capture to promote learning and transfer of knowledge. *American Journal of Occupational Therapy, 71*(3), 1–7. <https://doi.org/10.5014/ajot.2017.023374>
- Halstead, J. A. (2013). The NLN's fair testing imperative and implications for faculty development. *Nursing Education Perspectives, 34*(2), 72–72.

- Hanna, K., Roberts, T., & Hurley, S. (2016). Collaborative testing as NCLEX enrichment. *Nurse Educator, 41*(4), 171–174. <https://doi.org/10.1097/NNE.0000000000000241>
- Hanshaw, L. G. (2012). Qualitative aspects of group-only testing. *College Student Journal, 46*(2), 419–426.
- Hanson, M. J. S., & Carpenter, D. R. (2011). Integrating cooperative learning into classroom testing: Implications for nursing education and practice. *Nursing Education Perspectives, 32*(4), 270–273. <https://doi.org/10.5480/1536-5026-32.4.270>
- Harrison, C., & Wass, V. (2016). The challenge of changing to an assessment for learning culture. *Medical Education, 50*(7), 704–706. <https://doi.org/10.1111/medu.13058>
- Heglund, S., & Wink, D. (2011). Impact of double testing on student knowledge in a professional issues course. *Journal of Nursing Education, 50*(5), 278–280.
<http://dx.doi.org.cupdx.idm.oclc.org/10.3928/01484834-20110131-06>
- Hickey, B. L. (2006). Lessons learned from collaborative testing. *Nurse Educator, 31*(2), 88–91.
- Institute of Medicine. (2010). *The future of nursing* (Report Brief). Retrieved from <http://nacns.org/wp-content/uploads/2016/11/5-IOM-Report.pdf>
- Iwamoto, D. H., Hargis, J., Taitano, E. J., & Vuong, K. (2017). Analyzing the efficacy of the testing effect using Kahoot™ on student performance. *Turkish Online Journal of Distance Education, 18*(2), 80–93.
- Jang, H., Lasry, N., Miller, K., & Mazur, E. (2017). Collaborative exams: Cheating? Or learning? *American Journal of Physics, 85*(3), 223–227.
<https://doi.org/10.1119/1.4974744>
- Jefferys, D. T., Hodges, S., & Trueman, M. S. (2017). A strategy for success on the National Council Licensure Examination for at-risk nursing students in historically black colleges

- and universities: A pilot study. *International Journal of Caring Sciences*, *10*(3), 1705–1709.
- Kantar, L. D. (2014). Assessment and instruction to promote higher order thinking in nursing students. *Nurse Education Today*, *34*(5), 789–794.
<https://doi.org/10.1016/j.nedt.2013.08.013>
- Killingsworth, E., Kimble, L. P., & 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. <https://doi.org/10.5480/14-1492>
- Knowles, M. S., Elwood III, F. H., & Swanson, R. A. (2015). *The adult learner: The definitive classic in adult education and human resource development* (8th ed.). New York, NY: Routledge.
- Lang, J. M. (2014, April 23). Making it stick. *The Chronicle of Higher Education*. Retrieved from <https://www.chronicle.com/article/Making-It-Stick/146143>
- Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education*, *47*(7), 674–682.
<https://doi.org/10.1111/medu.12141>
- Leight, H., Saunders, C., Calkins, R., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE-Life Sciences Education*, *11*(4), 392–401. <https://doi.org/10.1187/cbe.12-04-0048>
- LoGiudice, A. B., Pachai, A. A., & Kim, J. A. (2015). Testing together: When do students learn more through collaborative tests? *Scholarship of Teaching and Learning in Psychology*, *1*(4), 377–389. <http://dx.doi.org.cupdx.idm.oclc.org/10.1037/stl0000041>

- March, A. L., & Robinson, C. (2015). Assessment of high-stakes testing, hopeful thinking, and goal orientation among baccalaureate nursing students. *International Journal of Nursing Education Scholarship*, 12(1), 393–397.
<http://dx.doi.org.cupdx.idm.oclc.org/10.1515/ijnes-2014-0075>
- Martin, D., Friesen, E., & De Pau, A. (2014). Three heads are better than one: A mixed methods study examining collaborative versus traditional test-taking with nursing students. *Nurse Education Today*, 34(6), 971–977. <https://doi.org/10.1016/j.nedt.2014.01.004>
- Mbalamula, Y. S. (2018). Effect of group versus individual assessments on coursework among undergraduates in Tanzania: Implications for continuous assessments in universities. *Pedagogical Research*, 3(1), 1–8. <https://doi.org/10.20897/pr/85171>
- McClenny, T. (2016). *Student experiences of high-stakes testing for progression in one undergraduate nursing program* (Doctoral dissertation, University of West Georgia). Retrieved from <http://search.proquest.com/docview/1795584996>.
- Merriam, S. B. (2001). Andragogy and self-directed learning: Pillars of adult learning theory. *New Directions for Adult and Continuing Education*, 2001(89), 3–14.
<https://doi.org/10.1002/ace.3>
- Meseke, C. A., Nafziger, R., & Meseke, J. K. (2010). Student attitudes, satisfaction, and learning in a collaborative testing environment. *Journal of Chiropractic Education*, 24(1), 19–29.
- Mezirow, J. (1991). *Transformative dimensions of adult learning* (1st ed). San Francisco, CA: Jossey-Bass.
- Molsbee, C. P. (2013). Collaborative testing and mixed results. *Teaching and Learning in Nursing*, 8(1), 22–25. <https://doi.org/10.1016/j.teln.2012.09.001>

- Mueller-Joseph, L. J., & Nappo-Dattoma, L. (2013). Collaborative learning in pre-clinical dental hygiene education. *Journal of Dental Hygiene*, *87*(2), 64–72.
- Mulligan, N. W., & Peterson, D. J. (2015). The negative testing and negative generation effects are eliminated by delay. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1014–1025. <http://dx.doi.org/10.1037/xlm0000070>
- Nafziger, R., Meseke, J. K., & Meseke, C. A. (2011). Collaborative testing: The effect of group formation process on overall student performance. *Journal of Chiropractic Education*, *25*(1), 11–15.
- National Council of State Boards of Nursing. (2015). *NCLEX-RN examination: Detailed test plan for the national council licensure examination for registered nurses*. Retrieved from https://www.ncsbn.org/2016_RN_DetTestPlan_Educator.pdf
- National League for Nursing. (2012). *The fair testing imperative in nursing education* (Position Statement). Retrieved from http://www.nln.org/docs/default-source/about/nln-vision-series-%28position-statements%29/nlnvision_4.pdf?sfvrsn=4
- Oermann, M. H., & Gaberson, K. B. (2017). *Evaluation and testing in nursing education* (5th ed.). New York, NY: Springer.
- Pan, S. C., Pashler, H., Potter, Z. E., & Rickard, T. C. (2015). Testing enhances learning across a range of episodic memory abilities. *Journal of Memory and Language*, *83*, 53–61. <https://doi.org/10.1016/j.jml.2015.04.001>
- Pandey, C., & Kapitanoff, S. (2011). The influence of anxiety and quality of interaction on collaborative test performance. *Active Learning in Higher Education*, *12*(3), 163–174. <https://doi.org/10.1177/1469787411415077>

- Parsons, S. D., & Teel, V. (2013). Double testing: A student perspective. *Nursing Education Perspectives; New York*, 34(2), 127–128.
- Peck, S. D., Stehle Werner, J. L., & Raleigh, D. M. (2013). Improved class preparation and learning through immediate feedback in group testing for undergraduate nursing students. *Nursing Education Perspectives*, 34(6), 400–404. <https://doi.org/10.5480/11-507>
- Piper, K. (2016). Active learning: What do we know about It? *Australian Nursing & Midwifery Journal*, 24(3), 35–35.
- Pugh, D., & Regehr, G. (2016). Taking the sting out of assessment: is there a role for progress testing? *Medical Education*, 50(7), 721–729. <https://doi.org/10.1111/medu.12985>
- Racsmány, M., Szollosi, Á., & Bencze, D. (2018). Retrieval practice makes procedure from remembering: An automatization account of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(1), 157–166.
<http://dx.doi.org/10.1037/xlm0000423>
- Raupach, T., Andresen, J. C., Meyer, K., Strobel, L., Koziolk, M., Jung, W., ... Anders, S. (2016). Test-enhanced learning of clinical reasoning: A crossover randomized trial. *Medical Education*, 50(7), 711–720. <https://doi.org/10.1111/medu.13069>
- Retnowati, E., Ayres, P., & Sweller, J. (2017). Can collaborative learning improve the effectiveness of worked examples in learning mathematics? *Journal of Educational Psychology*, 109(5), 666–679.
- Rivaz, M., Momennasab, M., & Shokrollahi, P. (2015). Effect of collaborative testing on learning and retention of course content in nursing students. *Journal of Advances in Medical Education & Professionalism*, 3(4), 178–182.

- Rizzolo, M. A. (2015). The National League for Nursing project to explore the use of simulation for high-stakes assessment: Process, outcomes, and recommendations. *Nursing Education Perspectives, 36*(5), 299–303. <https://doi.org/10.5480/15-1639>
- Røykenes, K., Smith, K., & Larsen, T. M. B. (2014). ‘It is the situation that makes it difficult’: Experiences of nursing students faced with a high-stakes drug calculation test. *Nurse Education in Practice, 14*(4), 350–356. <https://doi.org/10.1016/j.nepr.2014.01.004>
- Sandahl, S. S. (2009). Collaborative testing as a learning strategy in nursing education: A review of the literature. *Nursing Education Perspectives, 30*(3), 171–175.
- Santo, L., Frander, E., & Hawkins, A. (2013). The use of standardized exit examinations in baccalaureate nursing education. *Nurse Educator, 38*(2), 81–84. <https://doi.org/10.1097/NNE.0b013e3182829c66>
- Schmidmaier, R., Ebersbach, R., Schiller, M., Hege, I., Holzer, M., & Fischer, M. R. (2011). Using electronic flashcards to promote learning in medical students: Retesting versus restudying. *Medical Education, 45*(11), 1101–1110. <https://doi.org/10.1111/j.1365-2923.2011.04043.x>
- Siegel, M. A., Roberts, T. M., Freyermuth, S. K., Witzig, S. B., & Izci, K. (2015). Aligning assessment to instruction: Collaborative group testing in large-enrollment science classes. *Journal of College Science Teaching, 44*(6), 74–82.
- Spiers, J. A., Williams, B., Gibson, B., Kabotoff, W., McIlwraith, D., Sculley, A., & Richard, E. (2014). Graduate nurses’ learning trajectories and experiences of problem based learning: A focused ethnography study. *International Journal of Nursing Studies, 51*(11), 1462–1471. <https://doi.org/10.1016/j.ijnurstu.2014.03.002>

- Spurlock, D. (2013). The promise and peril of high-stakes tests in nursing education. *Journal of Nursing Regulation, 4*(1), 4–8.
- Srougi, M. C., Miller, H. B., Witherow, S. D., & Carson, S. (2013). Assessment of a novel group-centered testing schema in an upper-level undergraduate molecular biotechnology course. *Biochemistry and Molecular Biology Education, 41*(4), 232–241.
- Stillwell, S. B., & Krautscheid, L. C. (2016). Answering student questions during examinations: A descriptive study of faculty beliefs. *Nursing Education Perspectives, 37*(3), 168–170.
- Stonecypher, K., & Willson, P. (2014). Academic policies and practices to deter cheating in nursing education. *Nursing Education Perspectives, 35*(3), 167–179.
<https://doi.org/10.5480/12-1028.1>
- Sullivan, D. (2014). A concept analysis of “high stakes testing.” *Nurse Educator, 39*(2), 72–76.
<https://doi.org/10.1097/NNE.0000000000000021>
- Tanriverdi, G., Ozyazicioglu, N., Atay, S., Sivrikaya, S. K., Gursoy, M. Y., & Cetin, A. (2017). The recommended solutions of nursing students to bridge the gap between the school and practice areas. *International Journal of Caring Sciences, 10*(1), 199–208.
- Tolsgaard, M. G., Kulasegaram, K. M., & Ringsted, C. V. (2016). Collaborative learning of clinical skills in health professions education: The why, how, when, and for whom. *Medical Education, 50*(1), 69–78. <https://doi.org/10.1111/medu.12814>
- Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition, 41*(3), 429–442. <https://doi.org/10.3758/s13421-012-0274-5>
- University of New Mexico-Gallup. (2017). *Snapshot of UNM-Gallup*. Retrieved from <http://www.gallup.unm.edu/main/about/snapshot.php>

- Vogler, J. S., & Robinson, D. H. (2016). Team-based testing improves individual learning. *Journal of Experimental Education, 84*(4), 787–803.
<https://doi.org/10.1080/00220973.2015.1134420>
- Warner, J. (2017). When standardized tests don't count. Retrieved June 20, 2017, from <https://www.insidehighered.com/blogs/just-visiting/when-standardized-tests-dont-count>
- Wiggs, C. M. (2011). Collaborative testing: Assessing teamwork and critical thinking behaviors in baccalaureate nursing students. *Nurse Education Today, 31*(3), 279–282.
<https://doi.org/10.1016/j.nedt.2010.10.027>
- Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology, 55*(1), 10–16.
<https://doi.org/10.1111/sjop.12093>
- Wissman, K. T., & Rawson, K. A. (2016). How do students implement collaborative testing in real-world contexts? *Memory, 24*(2), 223–239.
<https://doi.org/10.1080/09658211.2014.999792>
- Wood, C. M., & Ezebuihe, I. (2016). Bridging the theory practice gap using best practice assessment strategies. *Maryland Nurse, 18*(1), 3–5.
- Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(3), 485–492. <http://dx.doi.org/10.1037/xlm0000449>
- Zhang, N., & Henderson, C. N. R. (2017). Requiring students to justify answer changes during collaborative testing may be necessary for improved academic performance. *Journal of Chiropractic Education, 31*(2), 96–101. <https://doi.org/10.7899/JCE-16-5>

Zipp, J. F. (2007). Learning by exams: The impact of two-stage cooperative tests. *Teaching Sociology*, 35(1), 62–76.

Appendix A: IRB Approval

RE: Approval to use data

IRB Main Campus <irbmaincampus@xxx.edu>

Wed 5/17/2017 1:36 PM

To: June Eastridge <June.Eastridge@xxx.edu>; Hi June,

This study is not exempt. Exempt studies receive IRB review and are determined to be exempt. As you are not affiliated with XXX, the XXX IRB does not have oversight of the research and will not review the study. De-identified data sets do not meet the definition of human subjects and thus does not require IRB review. XXX IRB will not review the study and therefore will not give you a determination on letterhead. Let me know if you have more questions.

Thank you, XXXXX

XXXXXXXX XXXXX, XX, XXX Senior IRB Analyst, Office of the Institutional Review Board
The University of XXXXXXXXX XXX-XXX-XXXX or XXX-XXX-XXXX
IRBMainCampus@xxx.edu

The OIRB wishes everyone a great summer break! **The OIRB holds regular office hours and the XXX IRB reviews submissions throughout the summer.** You are still able to schedule consults and submit studies for review.



Appendix B: Statement of Original Work

I attest that:

1. I have read, understood, and complied with all aspects of the Concordia University-Portland Academic Integrity Policy during the development and writing of this dissertation.
2. Where information and/or materials from outside sources has been used in the production of this dissertation, all information and/or materials from outside sources has been properly referenced and all permissions required for use of the information and/or materials have been obtained, in accordance with research standards outlined in the *Publication Manual of The American Psychological Association*



Digital Signature

June A. Eastridge

Name

June 22, 2018

Date